

# THIS WEEK



## EDITORIALS

**WORLD VIEW** The researchers who warned last year about East African droughts **p.7**

**HIV TEST** New assay could speed screening in remote and rural areas **p.8**

**HEAR HEAR** Rainforest vine uses echo chamber to lure pollinating bats **p.8**

## Staying at home

*China's best young scientists are working abroad. The country is trying to change that — and when it succeeds, the results will be startling.*

The postdoctoral researcher is the workhorse of laboratories around the world, and principal investigators in countries ranging from the United States to Singapore have come to depend for their postdocs on well educated and hard-working researchers from China. Yet, ask senior scientists in China, especially bioscientists with experience of working overseas, and they'll say they want and need more postdocs back home — assuming that they have any at all.

Almost all the best young Chinese scientists with new PhDs choose to spend the formative period of their training for an independent career abroad. And those Chinese postdocs who do choose to work in China often desert academia for more lucrative careers in biotechnology, the pharmaceutical industry or contract research. The upshot is that, to get any work done, Chinese academic scientists have to keep training new groups of graduate students to fill the gap.

No doubt this is part of the reason that researchers in China, despite their reputation for hard work and putting in long hours, and despite the country's increasing investment in science, look extremely unproductive by international standards (see News Feature, page 22). One Chinese scientist described it as the ambitious nation shooting itself in the foot, twice — spending time and resources to train students and then sending them abroad to work for a competitor.

Yet the current system leaves talented people little choice but to leave. In particular, established laboratories overseas have more experience and success in producing good publications, which help to secure jobs for their postdocs. These laboratories are often better equipped than in China, and a postdoc job in the United States or Europe pays as well as or better than a more senior principal-investigator position in China.

And even while senior staff at Chinese universities and institutes despair of finding good-quality postdocs, they know that when a full-time post comes up, it will probably go to someone who has spent time abroad. A mediocre CV that includes stints in the United States or another foreign country, time and again trumps solid publications based on work done locally (see *Nature* **457**, 522; 2009). The lesson is clear — if you want to make it in China, go and train somewhere else.

But the tide may be turning. The Institute of Neuroscience in Shanghai (see page 22) together with institutions in Beijing such as the National Institute of Biological Sciences, the Institute of Biophysics, Tsinghua University and Peking University, are starting to form a critical mass — a domestic ring of excellence that offers the high-quality start to an independent career that students have up to now been forced to seek elsewhere.

Over the next few years, the effect could be hugely significant. With the US economy failing and some top institutes in China already considering ways to increase postdocs' salaries, benefits could start to converge. New facilities in China are as good as, or better than, in the West, and research in China offers other attractions, including easier access to primates. Research pastures at home will start to look as green as those abroad.

New recruits to the Chinese army of PhDs would do well to consider their country's own facilities as career launchpads. And Chinese institutions can do their part by looking seriously at domestically bred talent when recruiting, rather than running after returnees. Once that mentality starts to change, and Chinese laboratories fill up with four or five postdocs each, those productivity figures will shift rapidly. ■

## Safe, not secure

*A legal victory for US stem-cell research will not end the uncertainty over this disputed field.*

When US district court judge Royce Lamberth last week threw out a 2009 lawsuit against federally funded research on human embryonic stem cells, he concluded that Congress has acted reasonably in supporting the work. It matters little that Lamberth made clear he was only grudgingly falling into line with a higher court's decision in April — he was right.

This reading of the law does not, and will never, satisfy those who see the work as unethical. But six successive, elected Congresses have seen fit to fund the research, and what began as a US\$10-million investment by the National Institutes of Health (NIH) in 2002 has grown to an estimated \$125 million this year.

Lamberth's judgment (see page 14) significantly diminished the chances of his decision being reversed on appeal by higher courts. Scientists who take NIH funds for this research can breathe a large, if not total, sigh of relief when it comes to legal challenges.

Still, it must be acknowledged that the lawsuit — and the decade of politics around stem cells that preceded it — has inflicted serious damage on the field. From one year to the next, and sometimes one month or day to the next, stem-cell researchers have not been able to count on reasonable and steady government funding. Stem-cell scientists interviewed last week were notably cool on the new ruling. For many, it is irrelevant, because they have moved to politically safe and scientifically exciting work with induced pluripotent stem cells. Others have retreated to using adult stem cells.

It is unknown how many young people have shunned careers in the field because of the uncertainty, but is likely to be a considerable number. Those established embryonic stem-cell scientists who have funding from private or state grants continue with the work, but even now many of them consider NIH funding to be unreliable given its volatile political profile. The law has got this one right, but being right is rarely enough. ■



## We thought trouble was coming

**Chris Funk** explains how his group last year forecast the drought in Somalia that is now turning into famine — and how that warning wasn't enough.

Not since a million people died in Ethiopia and Sudan in 1984 and 1985 has the world seen such a potential for famine as it does now, with food emergencies occurring in Somalia, Ethiopia and Kenya. But although the ongoing disaster in East Africa is dire, it was not unexpected. In fact, I am part of a group of scientists that successfully forecast the droughts behind the present crisis.

I work with the Famine Early Warning Systems Network (FEWS NET), which was set up by the US Agency for International Development to help policy-makers prevent such humanitarian disasters. The network identifies where food aid is needed by the most food-insecure populations of the developing world, whose livelihoods are tied to rain-fed subsistence agriculture and pastoralism.

Last summer, our group was meeting when a La Niña weather system was forecast. We knew that such an event could bring trouble, and we issued an alert that East Africa might experience severe droughts.

We based this conclusion on information from three sources. First, we knew that La Niña events are commonly associated with weakened rains in the Horn of Africa from October to December.

Second, from work on the ground, we knew that persistent poor rains at the end of the past decade, combined with high food prices, had weakened the population's resilience to food emergencies.

And third, research has linked warming in the Indian Ocean as a result of climate change to drying of March-to-June rains in East Africa. This warming has intensified the negative impact of La Niña events; it was the chance that both the autumn and spring rainy seasons could be affected, back to back, that really concerned us.

Sure enough, the autumn 2010 rains were poor, or failed completely. The outlook for famine or survival then rested on the spring rains. April came without rain. May came without rain. And we feared the worst.

The situation on the ground quickly deteriorated. FEWS NET runs a food-price tracking system that showed that the price of maize (corn) in Kitui, Kenya, had soared by 246% in 12 months. And the value of a goat in Bardera, Somalia, usually sold to buy grain, had halved. Satellite measurements of vegetation health tracked the emerging drought in disturbing detail. FEWS NET put out a second alert on 7 June that warned: "This is the most severe food security emergency in the world today, and the current humanitarian response is inadequate."

Two months on, the grim statistics show that the massive crisis is outstripping the international resources available to address it. Famine conditions are expected to spread farther across Somalia, and large areas of Kenya and Ethiopia could see food availability fall to crisis levels. In all, some 11.5 million people across East Africa need emergency assistance.

So what went wrong? Why weren't the warnings — before and during the drought — enough to avert a food crisis that might turn into famine? Much of the problem is tied to political issues, especially in Somalia, but there are also strong climate and agricultural components.

The global climate models used by the Intergovernmental Panel on Climate Change were never intended to provide rainfall trend projections for every region. These models say that East Africa will become wetter, yet observations show substantial declines in spring rainfall in recent years. Despite this, several agencies are building long-term plans on the basis of the forecast of wetter conditions. This could lead to agricultural development and expansion in areas that will become drier. More climate science based on regional observations could be helpful in addressing these challenges.

On the agricultural side, crop yields are very low, and not increasing much. The amount of farmland per person is decreasing rapidly, as population growth places more pressure on a limited landscape. Agricultural progress is not keeping pace with population growth.

Declines in agricultural capacity are exacerbated by warming of the Indian Ocean, which is reducing the onshore flow of moisture during the spring rainy season, creating more frequent droughts. These trends are having an impact in southern Ethiopia, central and eastern Kenya and southern Somalia — those regions that have been hardest hit this year. Warmer and drier weather is shrinking the amount of land that is suitable for farming, leaving burgeoning food-insecure populations exposed to more frequent and severe climate shocks.

Emergencies such as the one in East Africa will become more common unless there is a focus on improving agricultural production. Ironically, the fact that crop yields are low creates a tremendous opportunity for improvement. A 50%, or even 100%, increase in yields is feasible, and would greatly improve the availability of food. Better integrated markets and increased grain and water storage can help to keep resources on hand for lean times. In the long term, a more resilient system is needed, rather than an increase in the number of emergency grain shipments. Then, when disaster strikes, surplus food can be moved around the region — from Tanzania to Somalia, say.

Better regional climate-change and forecast models, combined with more effective agriculture in drought-threatened areas will not solve all problems, but they should reduce the need for emergency responses, and make such measures more effective when they are necessary. ■

**Chris Funk** is in the Climate Hazard Group, part of the Geography Department at the University of California, Santa Barbara, USA.  
e-mail: cfunk@usgs.gov

APRIL CAME  
WITHOUT RAIN.  
MAY CAME  
WITHOUT RAIN.  
AND WE  
FEARED THE  
WORST.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/uo7at](http://go.nature.com/uo7at)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## EVOLUTIONARY BIOLOGY

### Chimp brains don't shrink

The human brain shrinks with age in what seems to be an evolutionarily new phenomenon, report Chet Sherwood of the George Washington University in Washington DC and his colleagues. They found that no parallel reduction in brain size seems to occur in our closest relative, the chimpanzee.

The researchers used magnetic resonance imaging to measure the volumes of whole brains, as well as regions of the frontal lobe and hippocampus, in 87 humans aged 22–88 and 99 chimpanzees aged 10–51. These volumes all fell with age in humans, but maintained a stable size in chimps. The team speculates that the shrinkage occurs because, compared with chimps, humans have evolved an extended lifespan, which amplifies normal cellular ageing processes.

*Proc. Natl Acad. Sci. USA*  
<http://dx.doi.org/10.1073/pnas.1016709108> (2011)

## PALAEONTOLOGY

### A trilobite's footprint

A 500-million-year-old trilobite fossil shows the marine arthropod literally caught in its tracks.

Palaeontologists scouring a Middle Cambrian-era outcrop west of Prague in the Czech Republic have discovered a rock bearing a 3-centimetre-long path that leads to the fossil of a small trilobite, *Agraulos*



*ceticephalus* (pictured). Oldřich Fatka of Charles University in Prague and his colleague Michal Szabad say the tracks may represent debris left over from a systematic excavation of the sea floor by the creature, and show this trilobite's final meal. If so, the fossil indicates that these animals were low on the food chain, sifting through small particles at the bottom of the sea.  
*Palaeontol. Z.* <http://dx.doi.org/10.1007/s12542-011-0102-4> (2011)



## PLANT SCIENCE

### A leaf that's loud and proud

Many plants lure pollinators to their flowers with diverse colours and patterns, but *Marcgravia evenia* (pictured) has evolved to attract pollinators that rely on sound rather than sight. The Cuban rainforest vine grows a deep cup-shaped leaf above its flowers that creates a distinct echo for nectar-feeding bats.

Ralph Simon at the University of Ulm in Germany and his colleagues analysed the leaf's acoustic properties and found that its unique shape produces a strong, constant echo across a

range of sound-source angles. They then trained bats to seek a feeder hidden in artificial foliage. The animals found feeders topped with the cup shape in an average of 12 seconds — around half the time it took them to locate unadorned feeders or those under other leaf shapes.

The team concludes that dish-shaped leaves help this rare plant to attract a key pollinator, as well as reducing bats' energy expenditure.

*Science* 333, 631–633 (2011)  
See [go.nature.com/cdqcja](http://go.nature.com/cdqcja) for more on this story.

## MEDICINE

### Rapid HIV test for remote areas

A microfluidic device could assist with the challenge of detecting infectious agents such as HIV in remote areas of the world.

Samuel Sia at Columbia University in New York and his colleagues developed a low-cost assay that can identify HIV and syphilis using just 1 microlitre of blood. They tested their 'mChip' assay in Rwanda on more than 200 samples and found that the test was faster than its lab-based counterpart, but offered

equivalent sensitivity and specificity.

The chip could enable rapid diagnosis of life-threatening infectious diseases in isolated locations, and provide important pre-screening of blood donations, the team says.

*Nature Med.* <http://dx.doi.org/10.1038/nm.2408> (2011)

## GENOMICS

### Breast-cancer weaknesses found

The blockade of highly active gene variants associated with breast cancer, such as *HER2*, is a standard approach to combating the disease. But

R. KLINGHOLZ/AGENTUR FOCUS/EYEVINE



proteins that are not mutated or overexpressed may also be crucial to a cancer cell's ability to survive.

Christopher Lord and his team at the Institute of Cancer Research in London used RNA interference screening — which silences individual genes by interfering with protein synthesis — on a panel of 20 breast-cancer cell lines. These lines represented all the major subtypes of the disease, offering a chance to find tumour-specific vulnerabilities and potential drug targets.

The authors uncovered a series of novel Achilles heels for particular tumour cell lines, including the reliance of cells deficient in the protein PTEN on TTK kinase, and the sensitivity of oestrogen-receptor-positive breast-cancer cells to the loss of the kinase ADCK2.

*Cancer Discov.* <http://dx.doi.org/10.1158/2159-8290.CD-11-0107> (2011)

## MATERIALS

## Butterfly wings turned to metal

The tiny scales that lend colour and texture to a butterfly's wings can be used as organic templates to make metallic nanostructures (**pictured**).

Jiajun Gu, Di Zhang and their colleagues at Shanghai Jiao Tong University in China first added gold nanoparticles to the surface of the wing scales, which are made from the tough polysaccharide chitin. They then deposited a metallic coating on the structures before removing the organic material. The large variety of butterfly-wing morphologies means that every butterfly bears scales of different shapes and size.

The authors' fabrication method is adaptable to at least

seven common metals. With almost 175,000 species of butterfly and moth to choose from, materials scientists could generate a wealth of intact, three-dimensional shapes with submicrometre resolution. The nanostructures have many potential applications — for example, in the design of photonic crystals. *Angew. Chem. Int. Edn* <http://dx.doi.org/10.1002/anie.201103505> (2011)

## ASTROPHYSICS

## Welcome to the neighbourhood

The Sun's neighbourhood is a little more crowded than we realized. Astronomers have discovered two ultracool brown dwarfs around 5 parsecs away — just four times farther than the distance to our next-nearest star, Proxima Centauri.

Brown dwarfs are objects bigger than Jupiter, but not big enough for their gravity to ignite fusion, as in a star. Ralf-Dieter Scholz and his colleagues at the Leibniz Institute for Astrophysics in Potsdam, Germany, found the brown dwarfs using NASA's Wide-field Infrared Survey Explorer. They looked for objects that had moved a lot in the sky since two previous ground-based surveys, indicating that these objects are close to us.

*Astron. Astrophys.* 532, L5 (2011)

## IMMUNOLOGY

## An antibody for all flu strains

A person who contracted the strain of influenza responsible for the 2009–10 pandemic produced an antibody with the remarkable ability to block all strains of influenza A.

## COMMUNITY CHOICE

The most viewed papers in science

## CHEMISTRY

## Iron nanoparticles into blood

**HIGHLY READ**  
on [pubs.acs.org](http://pubs.acs.org)  
in July

Tiny particles of iron oxide can have their surface chemistry manipulated to make them soluble in water and so more useful in medical applications. Iron oxide

nanoparticles have potential for drug delivery and as contrast agents for magnetic resonance imaging, but they are generally made in organic solvents, and keeping them stable and soluble in the body's aqueous environment has proved difficult.

Yuping Bao and her colleagues at the University of Alabama in Tuscaloosa found a way to swap the chemical groups originally coating the nanoparticles for groups that make the particles soluble in biological solvents. The process worked best when the original groups were triethylphosphine oxide and their replacements were poly(acrylic acid), polyethylenimine or glutathione — provided at a ratio of five to each nanoparticle.

*Langmuir* 27, 8990–8997 (2011)

## ECOLOGY

## Pollination drives floral clock

The closure of flower heads is driven not only by falling light or temperature, as previously thought, but also by pollinators. In a set of field experiments, Jochen Fründ at Georg August University in Göttingen, Germany, and his team show that flowers in the Asteraceae family, including smooth hawkbeard (*Crepis capillaris*), close within three hours of pollination. By contrast, the heads of unpollinated flowers stayed open for hours longer — until late afternoon.

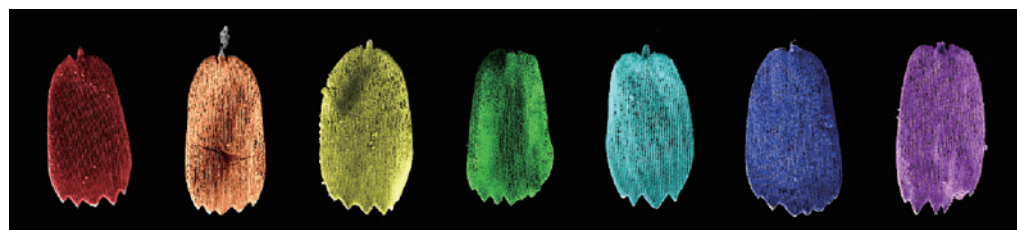
Flower-head closure earlier in the day affects interactions between plants and the pollinators of the surrounding area, and can change the times at which some insects are active. The authors suggest that their results be considered in plant and pollinator field surveys to ensure accuracy.

*Ecol. Lett.* <http://dx.doi.org/10.1111/j.1461-0248.2011.01654.x> (2011)

## NATURE.COM

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)





# SEVEN DAYS

The news in brief

## POLICY

### Stem-cell lawsuit

To the relief of stem-cell scientists, a District of Columbia judge has thrown out a lawsuit that sought to block US government funding of research using human embryonic stem cells. The decision, released on 27 July, came 11 months after the same judge issued a preliminary injunction on such research. See page 14 for more.

### Primate research

Research using primates should continue in the United Kingdom because of its importance to human health and basic science, a report released on 27 July concluded. But the review, commissioned by funding bodies the Medical Research Council, the Wellcome Trust and the Biotechnology and Biological Sciences Research Council, found that 9% of projects using non-human primates between 1997 and 2006 showed no obvious scientific or medical benefit. Around 3,000 primates — mostly macaques and marmosets — are used in UK labs each year. See [go.nature.com/6p6sad](http://go.nature.com/6p6sad) for more.

### Yucca's successor

To reinvigorate its quest for a permanent storage site for nuclear waste, the United States needs an independent waste-management organization with dedicated funding, says an interim report by a White House commission, published on 29 July. The report also calls for a public-engagement process to identify a viable replacement for the proposed Yucca Mountain repository in Nevada, which remains in legal limbo following the decision by President Barack Obama's administration to shutter the project. The commission plans to release its final report in



ESO/S. STANGHELLINI

## ALMA ready to study the Universe

A 16th radio antenna (pictured) was delivered last week to the Atacama Large Millimeter/submillimeter Array (ALMA), 5,000 metres above sea level in Chile's Atacama Desert. This milestone means that the radio telescope can begin original science observations in September. ALMA, an international

collaboration with a cost well over US\$1 billion, will study the Universe at submillimetre wavelengths, providing a new window on the cool star-forming regions of the Milky Way and on early galaxies. Its complete array (54 12-metre-diameter antennas, and 12 7-metre antennas) won't be in place until 2013.

January. See [go.nature.com/k2piu4](http://go.nature.com/k2piu4) for more.

### Research integrity

Britain's oversight of research integrity is "unsatisfactory", according to a 28 July report from the House of Commons Science and Technology Committee, a cross-party group of politicians. The report repeats recommendations made last year by the UK Research Integrity Futures Working Group to set up an independent advisory body on the subject, and calls for an external regulator for research integrity.

### Fuel standards

US President Barack Obama announced new fuel economy standards on 29 July, unveiling an agreement with car

makers that would bring the average fuel efficiency for all cars and trucks sold to 23.2 kilometres per litre (54.5 miles per US gallon) by 2025, nearly double the current average. According to preliminary estimates by the International Council on Clean Transportation in Washington DC, that would bring emissions down to around 107 grams of carbon dioxide per kilometre travelled — still more than the target of 95 g CO<sub>2</sub> per km set by the European Commission for 2020.

officer of the US National Academy of Sciences, as its new science adviser. Colglazier's appointment was announced on 25 July; he succeeds geneticist Nina Fedoroff, who ended her three-year stint in the post last July.

### Marburger dies

John 'Jack' Marburger, who served as science adviser to US President George W. Bush for eight years, died on 28 July, aged 70. A physicist with a background in lasers and nonlinear optics, Marburger (pictured) was president of Stony Brook University in New York from 1980 to 1994, then in 1998 became director of Brookhaven National Laboratory in Upton, New

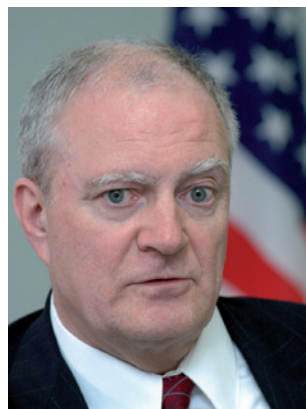
## PEOPLE

### US science adviser

The US Department of State has appointed physicist William Colglazier, who recently retired as executive

York. As science adviser during 2001–09, he staunchly defended Bush's policies in the face of angry criticism from researchers.

E. ALTERMAN/BLOOMBERG VIA GETTY IMAGES



## NASA loses leaders

NASA, an agency seemingly uncertain of its mission, is losing its most senior power couple. Laurie Leshin, deputy associate administrator of NASA's exploration division, and her husband Jon Morse, head of astrophysics in the science division, are leaving NASA for the Rensselaer Polytechnic Institute (RPI) in Troy, New York. Morse, who has overseen the astrophysics division through its toughest times, will become the RPI's associate vice-president for research, the university announced on 27 July. Leshin will become dean of the school of science. Both start at the beginning of October.

## RESEARCH

### China's deep dive

China's *Jiaolong* submersible completed a test dive to 5,188 metres in the central Pacific Ocean on 28 July — well past its previous 3,759-metre record and a depth that, in theory, gives China access to more than 70% of the ocean floor. The 8-metre long, 22-tonne craft is designed to dive to 7,000 metres. Japan's *Shinkai* submersible is rated to 6,500 metres, and Russia's *Mir* to around 6,000 metres. The venerable US submersible *Alvin* is being upgraded from its current 4,500-metre depth rating to 6,500 metres. See [go.nature.com/doysot](http://go.nature.com/doysot) for more.

### Climate data freed

After years of bitter dispute with climate-change sceptics, the Climatic Research Unit (CRU) at the University of East Anglia in Norwich, UK, has released most of the raw meteorological data it used to put together a data set of global land temperatures. CRU had resisted releasing the data under freedom-of-information requests, but was ordered to do so last month by the UK Information Commissioner. Only data from 19 stations in Poland are not included. The Polish meteorological service, say CRU officials, refused to have

its data publicly released. See [go.nature.com/tyi2ty](http://go.nature.com/tyi2ty) for more.

### Genome diversity

The consumer genetics firm 23andMe announced plans on 26 July to offer free genome scans to 10,000 African Americans in an effort to boost the ethnic diversity of its customers. The company, based in Mountain View, California, also hopes to determine whether genetic variants linked to diseases in other ethnic groups apply to African Americans. Ninety-six per cent of all such genome-wide association studies have been conducted in people of European ancestry (see *Nature* 475, 163–165; 2011), and the majority of 23andMe's current customers are also of European descent.

## BUSINESS

### Gene patents upheld

The US biotechnology industry's ability to patent genes, threatened by a lawsuit in which seven such patents were ruled invalid, seems safe again. Last March, a federal judge had ruled that patents for genes associated with breast and ovarian cancer were improperly granted, because they referred to isolated DNA that was a product of nature, and was

## COMING UP

### 5 AUGUST

NASA's Juno mission is scheduled to launch for Jupiter. See page 13 for more.

[www.nasa.gov/juno](http://www.nasa.gov/juno).

### 7–12 AUGUST

The Ecological Society of America has its annual meeting in Austin, Texas. The theme is preserving and enhancing a sense of "Earth stewardship".

[www.esa.org/austin](http://www.esa.org/austin)

therefore not patentable under US law. But on 29 July, a New York appeals court said that isolated DNA molecules were sufficiently distinct from their natural forms to be eligible for patent protection. The patents are held by Myriad Genetics, based in Salt Lake City, Utah; the issue is likely to require further judgments in higher courts. See [go.nature.com/lhmdem](http://go.nature.com/lhmdem) for more.

### RNAi shutdown

Pharmaceutical powerhouse Merck is shutting down the San Francisco research facility it acquired in 2006 when it paid US\$1.1 billion for Sirna Therapeutics, a California biotechnology firm specializing in RNA interference (RNAi) as a treatment. The 29 July announcement seems to be another blow to the RNAi field, which has been hit hard by pharma restructurings in the past year (see *Nature* 468, 487; 2010).

Separately, Merck, which is headquartered in Whitehouse Station, New Jersey, announced job cuts of 12–13% by 2015, on top of the 17% cuts it announced after purchasing Schering-Plough in November 2009. See [go.nature.com/av5uho](http://go.nature.com/av5uho) for more.

► **NATURE.COM**

For daily news updates see:

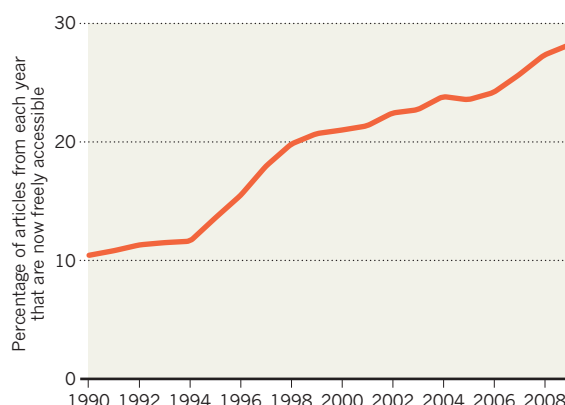
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

The 19 July indictment of activist and ethics researcher Aaron Swartz, who downloaded more than 4 million articles from the academic journal archive JSTOR, unleashed a ferment of support for the principle that research papers should be freely available. Unremarked so far is that the proportion of research papers freely available is increasing (see chart) thanks to public-access mandates by governments and funding agencies, and the success of open-access publishers. See [go.nature.com/gmpdbq](http://go.nature.com/gmpdbq) for more.

### FREEING RESEARCH ARTICLES

More than one-quarter of biomedical research articles indexed in PubMed in recent years are freely accessible.



SOURCE: D. LIPMAN/NCBI

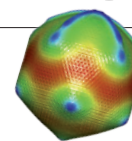


# NEWS IN FOCUS

**MEDICINE** After a victory, stem-cell researchers worry about the next battle **p.14**

**POLICY** Suspended polar-bear researcher becomes embroiled in political row **p.16**

**PUBLISHING** Microsoft and Google offer free metrics for all **p.18**



**BIOLOGY** Do giant viruses belong to a fourth domain of life? **p.20**

NASA/JPL-CALTECH



The Juno spacecraft will probe the Jovian atmosphere for water and look for signs of a solid core.

## PLANETARY SCIENCE

# Closing in on Jupiter's past

*NASA's Juno mission aims to reveal how the Solar System's largest planet was formed.*

BY ERIC HAND

It will be one wild ride. After arcing over Jupiter's north pole, the spacecraft will rip past the planet's equator at 60 kilometres per second, threading the gap between swirling cloud tops and a zone of high-energy radiation that would fry its delicate electronics. It will then swing out into space and repeat the harrowing journey 32 times more.

That's the plan for Juno, a US\$1.1-billion NASA mission that is due to launch on 5 August or soon after. If all goes well, five

years from now the spacecraft will drop into the highly elliptical orbit, which at closest approach will allow it to probe Jupiter's murky depths. The polar orbit will limit Juno's exposure to the worst of the radiation belt encircling Jupiter's equator, where the planet's magnetic field whips up electrons to nearly the speed of light. It also gives the spacecraft a Janus-like view. Instruments aimed away from Jupiter will map the radiation belts and magnetic fields, whereas those trained on the planet will probe its opaque layers for chemical and gravitational clues to its origins.

Crucial in that search will be an inventory of oxygen — sequestered as water vapour in Jupiter's atmosphere — and what it says about where and when the planet formed. Because Jupiter is likely to have formed first among the planets and because its powerful gravity has held its initial ingredients in place, Juno's results will also have broader significance. "Understanding the history of water across the early Solar System is a fundamental question, and Jupiter is going to give you the first clue," says Scott Bolton of the Southwest Research Institute in San Antonio, Texas, ►



► and principal investigator on the mission.

In 1995, a probe dropped by the Galileo mission found many volatile elements, such as nitrogen and argon, in higher proportions than expected at Jupiter's distance from the Sun. This suggested that Jupiter either migrated to its current location after forming elsewhere, or that it incorporated many comet-like building blocks from the Solar System's colder reaches. But because the probe descended through a rare dry spot with little water vapour, Galileo could not get a global read for Jupiter's oxygen. That left "a big hole" in what researchers know about the planet, says Tobias Owen at the University of Hawaii in Hilo, an investigator on the Galileo and Juno missions. This time, Juno will try to measure water content by detecting microwaves emitted by Jupiter's atmosphere. The amount of water present at different depths in the atmosphere alters the strength of the emission at different frequencies.

If Jupiter proves to be as enriched in oxygen as it is in other volatiles, that could lend support to a colder, more distant origin. Alternatively, the presence of even more oxygen would bolster models proposing that Jupiter formed close to its present orbit, with water ice trapping other volatiles. And if the global oxygen abundance is as low as that found by Galileo, "then we really have to open ourselves

up to new ideas," says Bolton.

Another key Juno experiment will try to identify whether Jupiter has a core — the roughly ten Earth masses of ice and rock that many theorists say would have been necessary to allow the runaway accretion of the hydrogen and helium gases that make up most of the planet. The experiment will look for the subtle effect that a core's gravitational pull would have on the flight of the spacecraft.

**"Understanding the history of water across the early Solar System is a fundamental question."**

But Alan Boss, a theorist at the Carnegie Institution for Science in Washington DC, says that the presence or absence of a core won't determine Jupiter's origin conclusively. There is an alternative formation model, called disk instability, in which a perturbation in a thick cloud of gas can cause it to shrink rapidly to form a giant planet — and this model works with or without a core. Moreover, Boss says, Jupiter's core could have changed over time. And, he says, laboratories on Earth are just beginning to understand the behaviour of the highly compressed hydrogen that makes up the bulk of Jupiter's interior, and which matters most in understanding its

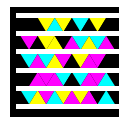
structure. "To claim that Juno will solve the question of Jupiter's formation doesn't seem to be supported by what we know right now," he says. But Bolton says that the data will certainly help to constrain the theorists. "Slowly, you can put the handcuffs on these guys," he says.

All the debate assumes that Juno will survive long enough to get the data. Engineers have tried to protect it from the intense radiation, placing instruments behind elaborate periscope-like mirrors, and putting the most vulnerable electronics into a titanium box known as 'the vault'. Even so, Juno's design limits it to just 33 polar orbits, one every 11 days, before it is sent plunging into Jupiter, to avoid the risk of its hitting the moon Europa and possibly contaminating it with terrestrial microbes.

Within those precious orbits, Bolton

hopes that his team can begin to understand how Jupiter was made — a question that has grown more relevant with the discovery of many Jupiter-mass planets in distant solar systems. "Jupiter is our archetype," Bolton says. "It's the only one we have." ■

**NATURE.COM**  
Take a video tour of Juno's mission:  
[go.nature.com/sosnyj](http://go.nature.com/sosnyj)



Scan the tag above with the free app from **gettag.mobi**

## REGENERATIVE MEDICINE

# Court quashes stem-cell lawsuit

*US judge throws out case meant to halt federal funding, but research remains vulnerable.*

BY MEREDITH WADMAN

**W**as the case a fluke or a forewarning? Now that a federal judge has thrown out a lawsuit that sought to halt US government funding of research using human embryonic stem cells, scientists who depend on that support are left wondering whether the battle is truly over, or is merely moving on to a different arena.

Chief Judge Royce Lamberth of the US District Court for the District of Columbia issued his decision on 27 July, acknowledging a higher court's opinion that overruled a preliminary injunction that he had placed to suspend the funding last August (see 'Trying times'). That injunction was in effect for only 17 days, but it threw the stem-cell research community into turmoil as hundreds of scientists faced a funding cut-off. All research on human embryonic stem cells at the US National Institutes of Health (NIH) in Bethesda, Maryland, was shut down and reviews of grant applications were left in limbo.

Last week's ruling decisively affirms the

government's legal ability to fund research on human embryonic stem cells. It is a significant blow to the plaintiffs in the case — James Sherley, a biological engineer at the Boston Biomedical Research Institute in Massachusetts, and Theresa Deisher, who runs AVM Biotechnology in Seattle, Washington. Both work with adult stem cells, which are isolated from tissues or organs rather than embryos. Speaking

to *Nature* on the day of Lamberth's decision, Deisher said: "It seems to be a complete reversal from his ruling last summer, so obviously it's quite a surprise."

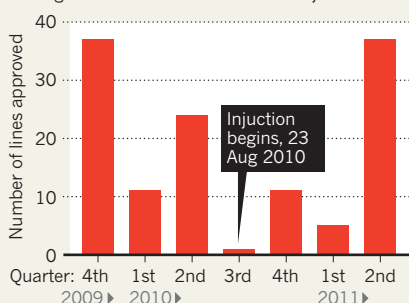
"We intend to review all of our options for appeal of this decision," says Steven Aden, a lawyer for the plaintiffs.

Legal experts say that any appeal by the plaintiffs will probably not be successful, given the strength of Lamberth's written opinion and the fact that the appeals court overturned the preliminary injunction. "They are likely to lose at every stage," says Alta Charo, who studies law and bioethics at the University of Wisconsin–Madison. She adds that the injunction may come to be seen as an anomaly, rather than as encouragement for those who aim to starve human embryonic stem-cell research of funding. "I think it's a one off," she says.

But for scientists who laboured under the threat of a funding shutdown for 11 months as the case unfolded, the outcome is bitter-sweet. Although NIH approval of new stem-cell lines has resumed, and even accelerated (see 'Bouncing back'), some say that it will

## BOUNCING BACK

The NIH has approved 42 stem-cell lines this year, ending a lull that followed the 2010 injunction.



take years to recover from the impact of the shutdown. Scientists who left the field in the interim might never return.

"Things have changed permanently. It's not just going to go back to the way it was — not immediately," says Meri Firpo, who researches stem-cell therapies at the University of Minnesota Stem Cell Institute in Minneapolis.

Candace Kerr, a stem-cell scientist at the Institute for Cell Engineering at Johns Hopkins University in Baltimore, Maryland, had NIH funding for human embryonic stem-cell work at the time of the shutdown, but has since moved entirely to work with induced pluripotent stem cells — adult cells that have been reprogrammed to act as stem cells. She has no plans for future studies on human embryonic stem cells. "The shock of what happened last year, coming out of nowhere, makes us really sceptical about moving forward," says Kerr.

Such scepticism may be warranted, given the history of political division over the scientific use of donated embryos in the United States. When issuing his injunction last year, Lamberth ruled that the NIH had violated a 1995 law, the Dickey–Wicker amendment, which forbids the government to fund "research in which a human embryo or embryos are destroyed". Three presidential administrations have interpreted the amendment to permit funding for research on lines of human embryonic stem cells, but not for the creation of such lines, which requires destruction of embryos. In 2001, US President George W. Bush restricted the funding to research involving a small number of existing lines. His successor, President Barack Obama, lifted this restriction in 2009, prompting the approval of dozens of new stem-cell lines; groups that oppose the research on moral grounds then launched a lawsuit to halt the funding. Sherley and Deisher were granted standing in the case after arguing that funding for human embryonic stem cells harmed their prospects of receiving NIH grants for their own research on adult stem cells.

The government appealed Lamberth's preliminary injunction, leading the appeals court first to place a temporary hold on the injunction and then, in April, to strike it down altogether, sending the case back to Lamberth's

court. Last week, constrained by the appeals-court decision, Lamberth rejected Sherley and Deisher's lawsuit. His written opinion attacked all of the plaintiffs' key arguments, finding that the Dickey–Wicker amendment's wording is "ambiguous" and that the NIH's interpretation that it allows funding of human embryonic stem-cell research is "reasonable", thus satisfying an administrative-law doctrine that courts must to defer to an agency's reasonable interpretation of an ambiguous law.

Hank Greely, director of the Center for Law and the Biosciences at Stanford University in California, says it is very likely that the appeals court will hold to the same position if the plaintiffs try to appeal the latest decision. "Then the only remaining hope for the plaintiffs is the Supreme Court," he says. The US Supreme Court accepts only about 1% of the thousands of cases it is asked to hear each year.

Nonetheless, human embryonic stem-cell research remains vulnerable. Charo notes that the current US Congress is so politically divided it is unlikely to enact a law either explicitly permitting or explicitly prohibiting government funding for the research. But, she says, "nothing in this decision and nothing in the Dickey–Wicker amendment" stops a new president from quashing research simply by refusing to fund it. Greely says that the best way to protect the research "is to get some real medical progress with stem cells" to prove the worth of the field.

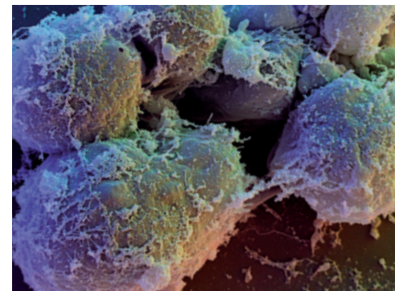
Kerr agrees. "The only way this debate's going to end is with a cure," she says.

The irony is that this will only happen if enough researchers commit to developing the field's potential — and they may be put off by the lawsuit's lingering shadow. "There has been one source of uncertainty after another," says Sean Morrison, outgoing director of the Center for Stem Cell Biology at the University of Michigan in Ann Arbor. In 2008, Morrison campaigned successfully to protect research involving human embryonic stem cells in the state's constitution, but in the past six months he has returned to earlier work with adult stem cells, which he will continue when he takes up a position at the University of Texas Southwestern Medical Center in Dallas this month.

"Eventually," says Morrison, "everyone starts looking around and saying, 'There's got to be easier ways to do science.'" ■ [SEE EDITORIAL P.5](#)

## TRYING TIMES

After an injunction against it, US courts upheld funding for stem-cell science.



D. SCHARF/SCIENCE FACTION/CORBIS

### 23 AUGUST 2010

Chief Judge Royce Lamberth of the US District Court for the District of Columbia issues an injunction suspending federal funding of research on human embryonic stem cells (pictured above).

### 9 SEPTEMBER 2010

The US Court of Appeals for the District of Columbia Circuit temporarily stays the Lamberth injunction.

### 6 DECEMBER 2010

The appeals court hears oral arguments for and against the preliminary injunction.

### 29 APRIL 2011

The appeals court permanently overturns the preliminary injunction.

### 27 JULY 2011

Lamberth (pictured, below), ruling on the merits, throws out the lawsuit. Plaintiffs' attorneys say they will "review all of our options for appeal".



AP PHOTO/J. SCOTT APPLEWHITE

### ➔ NATURE.COM

To read more about the stem-cell lawsuit, visit: [go.nature.com/ljonzx](http://go.nature.com/ljonzx)

## MORE ONLINE

### TOP STORY



Antibody that binds all influenza A strains could be blueprint for universal vaccine [go.nature.com/naecyt](http://go.nature.com/naecyt)

### OTHER NEWS

- How to design safer chemicals [go.nature.com/l9hzcy](http://go.nature.com/l9hzcy)
- Russian submersible *Mir* takes a dive into Lake Geneva [go.nature.com/kq1ezr](http://go.nature.com/kq1ezr)
- Survey of depression in China holds some surprises [go.nature.com/7q4iuk](http://go.nature.com/7q4iuk)

### ON THE BLOG



Science tries to look useful as US budget cuts loom [go.nature.com/asr5bo](http://go.nature.com/asr5bo)

FANCY





P. SOUDERS/CORBIS

A link between receding sea ice and polar-bear deaths has sparked political fallout.

## ENVIRONMENT

# Bear researcher frozen out

*Supporters question motives for suspension of government scientist.*

BY EUGENIE SAMUEL REICH

It was one of the most dramatic sightings ever made in an aerial survey of the Arctic: a dead polar bear, bloated like a gigantic beach ball, floating in open water north of the Beaufort Sea coastline in Alaska.

Researchers say that they spotted four dead polar bears during the survey, and surmised that the bears drowned in stormy waters as they searched for ever-receding sea ice. The idea that polar bears could drown like this became a rallying point for advocates of action on climate change, most notably appearing in former US vice-president Al Gore's film *An Inconvenient Truth* (2006).

Now, five years after the observations were reported, the bears have become the focus of charges ranging from scientific fraud to political interference in science. Last week, it emerged that the US Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) had suspended a researcher involved in the survey, wildlife biologist Charles Monnett. The reason, according to an 18 July memo from Monnett's supervisor, Jeffrey Loman, was an investigation into "integrity issues" by the Office of the Inspector General (OIG) at the US Department of the Interior, which oversees the BOEMRE. Climate-change sceptics were quick to jump on the news as evidence that the science of global warming had

been distorted. The BOEMRE has also halted a different polar-bear survey overseen by Monnett, pending further investigation.

Monnett's suspension was brought to light on 28 July by Public Employees for Environmental Responsibility (PEER), a watchdog group in Washington DC that is giving Monnett legal advice in the matter.

PEER released a transcript of an interview between criminal investigators at the OIG and Monnett, in which Monnett was told that he had been accused of scientific misconduct. He was then asked a series of questions relating to the paper in which he had reported the four

**Any accusation of playing into the oil-industry's hands is highly sensitive.**

drowned polar bears (C. Monnett and J. S. Gleason *Polar Biol.* **29**, 681–687; 2006), but was not told the specific allegations.

Jeff Ruch, executive director of PEER, says

that this does not conform with the Department of the Interior's scientific-integrity policy, which states that those accused of misconduct should be properly informed of the allegations against them, and that the allegations should be referred to a scientific-integrity official, not to criminal investigators. On 29 July, PEER filed a scientific and scholarly misconduct complaint against Monnett's superiors and the OIG, accusing them of violating the policy.

Ruch claims that the suspension is a politically motivated attack on Monnett's research at a time when the BOEMRE is considering whether to allow an expansion of oil drilling off Alaska's northern coast. The bureau denies this, and any accusation of playing into the oil industry's hands is highly sensitive, because the bureau (then known as the Minerals Management Service) was accused of poor oversight of the industry leading up to the Deepwater Horizon oil spill in the Gulf of Mexico in 2010. Ruch adds that Monnett is declining interviews because he has not been granted permission to do them by the bureau.

After a day of negative publicity generated by PEER's announcement, the bureau hit back. Spokeswoman Melissa Schwartz says that, contrary to the impression given by the transcript, Monnett's suspension was unrelated to scientific-integrity issues, his polar-bear finding or oil-drilling permits. She declined to say what it was related to.

But a 13 July memo to Monnett, provided to *Nature* by PEER, says that the investigation had uncovered information that raised concerns about his ability to act "in an impartial and objective manner" while handling a US\$1.1-million contract for a study of polar bears in the Canadian Arctic. A notice sent to Monnett by the OIG on 29 July further explained that although investigators may continue to query him on scientific integrity, they will now focus



on how Monnett awarded the research contract. This includes questions over whether Monnett complied with the Federal Acquisition Regulation, which is intended to ensure fair competition for US government contracts. The OIG adds that the inquiry is not criminal in nature, as the Department of Justice has already considered the case and declined to prosecute. Ruch says that Monnett's handling of the contract was transparent to his supervisors, and that his technical role meant he was not responsible for compliance with the regulation.

The project, begun in 2005, involves putting radio collars on polar bears found on the Canadian side of the Beaufort Sea, and tracking their position by satellite over several seasons. The study is funded by various sources, including the BOEMRE and the Canadian government. But on 13 July, the BOEMRE told scientists on the project to stop their work. The project's principal investigator, Andrew Derocher, a biologist at the University of Alberta in Edmonton, Canada, says he had no idea why. "To begin with, I thought it was related to budgetary issues in the United States. I've never seen anything like this in my life," he says.

Derocher says that data should continue to come in from collars until 2013, but the 'stop work' order may mean that he is unable to document his findings in a final report to the agency. Among those findings is that 2–4-year old polar bears tend not to stray far from their home range — the first time this age group has been tracked. This would mean that in the event of a large oil spill, bears that died from oil exposure would not be replaced quickly by bears from surrounding areas, says Derocher.

Drowned polar bears have not been reported by other scientists, but the hypothesis that a long search for sea ice makes it more likely that bears will get caught in stormy weather and drown is regarded as plausible. In January, scientists led by George Durner at the US Geological Survey in Anchorage, Alaska, reported the fate of an adult female bear as she swam more than 600 kilometres before reaching ice (G. M. Durner *et al.* *Polar Biol.* **34**, 975–984; 2011). When the researchers caught up with the animal, she had lost 22% of her body mass and her year-old cub.

This finding, corroborated by other studies, suggests that the major impact of receding sea ice on the bears is nutritional stress caused by a reduction of their hunting range, says Steven Amstrup, chief scientist at the campaigning organization Polar Bears International, headquartered in Bozeman, Montana, and a co-author of the study. But the observation that drowning can occur is important, he adds. "If this investigation is not about those observations then the BOEMRE owes it to him and to the public to say clearly what it is about." ■

## POLICY

# Conflict disclosure plan dropped

*The NIH will not require universities to create websites detailing researchers' financial ties.*

BY MEREDITH WADMAN

Francis Collins hailed it as a "new era of clarity and transparency in the management of financial conflicts of interest" (S. J. Rockey and F. S. Collins *J. Am. Med. Assoc.* **303**, 2400–2402; 2010). But the director of the US National Institutes of Health (NIH) may have spoken too soon when he described a new rule, proposed last year, that would require universities and medical schools to publicly disclose online any financial arrangements that they believe could unduly influence the work of their NIH-funded researchers.

*Nature* has learned that a cornerstone of that transparency drive — a series of publicly accessible websites detailing such financial conflicts — has now been dropped. "They have pulled the rug out from under this," says Sidney Wolfe, director of the Health Research Group at Public Citizen, a consumer-protection organization based in Washington DC. "It greatly diminishes the amount of vigilance that the public can exercise over financially conflicted research being funded by the NIH." It will also make it more difficult for "scholars to study the effects of conflicts of interest in universities", adds Sheldon Krinsky, who studies science ethics at Tufts University School of Medicine in Boston, Massachusetts.

The NIH's parent agency, the Department of Health and Human Services (DHHS), proposed the new rule in May 2010, after congressional and media investigations revealed that prominent NIH grant recipients had failed to tell their universities or medical schools about lucrative payments from companies that may have influenced their government-funded research. The DHHS called the proposed websites "an important and significant new requirement to ... underscore our commitment to fostering transparency, accountability, and public trust". Under the proposal, institutions with NIH-funded researchers would determine, grant by grant, if any financial conflicts existed for senior scientists on the grant. For example, these would include receiving consultancy fees, or holding shares in a company, "that could directly and significantly affect the design, conduct, or reporting" of the research. The institutions would post the details online, where they would stay for at least five years.

But a government official with knowledge of the ongoing negotiations on the rule says that the institutions will now be allowed to choose how to disclose this information, and will not be obliged to post it online. This is likely to make it much harder for members of the public to find these details, says Ned Feder, a senior staff scientist with the Project on Government Oversight. The watchdog group, based in Washington DC, wrote last month to the White House Office of Management and Budget (OMB) urging that the website requirement be protected. The OMB must sign off on the finalized form of the rule before it is published.

The OMB is also charged with enforcing an executive order issued by President Barack Obama in January. It requires government agencies to consider the costs of new regulations, and to tailor them to minimize cost and bureaucratic burden. "The websites don't appear out

of nowhere," says Heather Pierce, senior director of science policy at the Association of American Medical

Colleges (AAMC) in Washington DC. They would "require employees to not only create the website but to pull the information, review it, and make sure it is up to date and accurate".

That is not the only objection from the powerful academic lobbies. During the public comment period last summer, the Association of American Universities and the AAMC submitted a joint statement saying: "There are serious and reasonable concerns among our members that the Web posting will be of little practical value to the public and, without context for the information, could lead to confusion rather than clarity regarding financial conflicts of interest and how they are managed."

The two groups note that the Physician Payment Sunshine Act, a new law requiring drug firms to disclose their payments to physicians, requires the DHHS secretary to create a publicly available online database listing these payments. The groups suggest that the same model could work to publicize the financial conflicts of NIH-funded researchers.

Although the final rule is expected to be published soon, it is already long overdue. In May 2010, Collins said that it would be finalized by the end of that year. ■

# Computing giants launch free science metrics

*New Google and Microsoft services promise to democratize citation data.*

BY DECLAN BUTLER

Mapping the landscape of science is about to get easier than ever before. Google and Microsoft are rolling out free tools that will enable researchers to analyse citation statistics, visualize research networks and track the hottest research fields.

The systems could be attractive for scientists and institutions that are unable — or unwilling — to pay for existing metrics platforms, such as Thomson Reuters' Web of Knowledge and Elsevier's Scopus database.

Launched in 2004 as a search engine for academic publications, Google Scholar last month added Google Scholar Citations (GSC), which lets a researcher create a personal profile showing all their articles in the Google Scholar database ([go.nature.com/7wkpea](http://go.nature.com/7wkpea)). The profile also shows plots of the number of citations these papers have received over time, and other citation metrics including the popular *h*-index, which attempts to measure both the productivity of a scientist and the overall impact of their publications. The service is currently in invitation-only beta testing, but Google intends eventually to roll it out to all researchers.

Meanwhile, Microsoft Academic Search (MAS), which launched in 2009 and has a tool similar to GSC, has over the past few months added a suite of nifty new tools based on its citation metrics ([go.nature.com/u1ouut](http://go.nature.com/u1ouut)). These include visualizations of citation networks (see 'Mapping the structure of science'); publication trends; and rankings of the leading researchers in a field.

But although Microsoft's platform has many more features, Google Scholar has an enormous size advantage at present that makes its metrics far more accurate and reliable, say researchers. Google Scholar has indexed much more of the literature than has Microsoft, or indeed Web of Knowledge or Scopus. By contrast, MAS often turns up only a fraction of an author's true publications, which can result in its citation metrics having "absurdly low" values, says Péter Jacsó, an information scientist at the University of Hawaii in Honolulu.

"Microsoft Academic Search is still a nascent offering to the community," explains Lee Dirks, director of education and scholarly

communication at Microsoft Research Connections, the academic-collaboration arm of Microsoft Research. MAS's content surged from 15.7 million to 27.1 million publications between March and June, and that pace will continue, says Dirks. Anne-Wil Harzing at the University of Melbourne, Australia, who develops tools to extract citation metrics from Google Scholar, says that MAS has "great potential".

Some researchers question whether purely computational approaches can ever generate reliable bibliographic databases and citation

Scholar is so low that they do not greatly affect the accuracy of more robust metrics calculations such as the *h*-index.

Google Scholar also has an advantage over commercial providers in its extensive coverage of books — a significant research output in the social sciences and humanities — as well as conference proceedings, which are important outputs in the computing and engineering fields. Covering these is "crucial" to producing accurate metrics in these fields, says Ton van Raan, a bibliometrics expert at the Centre for Science and Technology Studies at Leiden

University, the Netherlands. Joel Hammond, director of product development at Thomson Reuters, points out that the Web of Knowledge already indexes conference proceedings, and that it plans to launch a book-citation index this autumn. Scopus has similar plans.

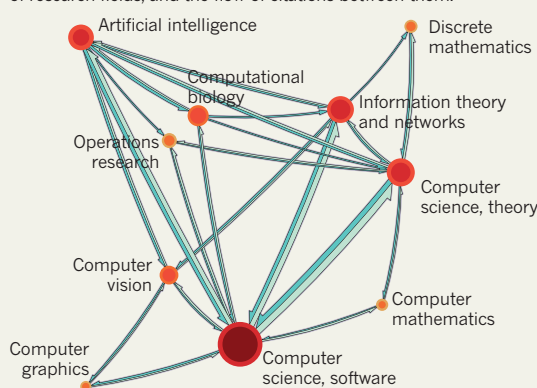
Neither MAS nor GSC see themselves as direct competitors of Web of Knowledge or Scopus, however. "This is not about competition, this is about providing an open platform for academic research," says Dirks. Acharya, who was born in India, says that he is driven by a humanitarian goal: making available to everybody services that were previously accessible only to those at richer institutions. He says he finds it "satisfying" that Google Scholar's server logs reveal widespread use by researchers in poorer countries, where commercial services are often unavailable.

Hammond says that Thomson Reuters controls which publications it indexes more strictly than do the free services, and argues that this makes its metrics calculations more reliable. Scopus takes a similar line. But others say GSC and MAS might eventually become good enough for many users. "They have the major advantage of being freely available to anyone, and with continued development I think they have the potential to become serious competitors to the commercial products," says Carl Bergstrom, a biologist at the University of Washington, Seattle, who collaborates with both Microsoft Research and Thomson Reuters to analyse citation data.

Van Raan agrees. "It is clear that the commercial citation index producers will be more and more in competition with these free-access facilities," he says. ■

## MAPPING THE STRUCTURE OF SCIENCE

Analysis of citations from 1.7 million computer-science publications in Microsoft Academic Search data reveals the relative importance of research fields, and the flow of citations between them.



metrics without some human intervention to clean up and check the data. Jacsó points out that the text-mining software used by MAS and GSC can sometimes extract erroneous bibliographic information from publications, for example by misidentifying author names or affiliations (P. Jacsó *Online Inform. Rev.* 34, 175–191; 2010).

Anurag Acharya, the Google engineer behind Google Scholar and its new metrics system, counters that it has long since dealt with such issues, and that a stack of recent improvements means that his system is working "better and better". Harzing adds that

critics often focus too much on such extreme bibliographic errors. She estimates that the overall level of errors in Google

**NATURE.COM**  
For more on  
metrics see:  
[www.nature.com/metrics](http://www.nature.com/metrics)





Groundwater loss threatens the coral-like formations of stromatolites in the Cuatro Ciénegas basin.

CONSERVATION

# Gene pool offers way to save Mexican oasis

*Commercializing genetic wealth will test biodiversity treaty.*

BY NICOLA JONES

**I**n the desert of northern Mexico, researchers are fighting to preserve an oasis of ancient life before it dries it up for good.

The 40-kilometre-long Cuatro Ciénegas basin (see map) is rich in springs, streams and pools, some of which have existed for tens of thousands of years. These host unique organisms, including living stromatolites — cyanobacterial colonies similar to some of the earliest life on Earth — and more than 70 endemic aquatic species, from snails to the world's only aquatic box turtle (*Terrapene coahuila*).

The Mexican government has designated almost 850 square kilometres of the area as a reserve, which prevents fishing but does not stop the extraction of groundwater or prevent some other forms of human activity. "Local people are still collecting firewood or plants for wax; there are horses mucking about in the ponds," says James Elser, a limnologist from Arizona State University in Tempe, who works in the region. "It hasn't been carefully managed."

Valeria Souza, a molecular biologist at the National Autonomous University of Mexico in Mexico City, is trying a new approach. In June, she wrangled a permit from the federal government allowing her to commercialize useful genes she finds in the region. Sharing

profits from patents with the local communities will, she hopes, encourage them to keep the water where it is, rather than extracting it for ranching and dairy farming.

In doing so, Souza is effectively testing how Mexico will operate under the Nagoya Protocol — a deal signed at the Convention on Biological Diversity summit in Nagoya, Japan, last October, that aims to regulate scientific access to genetic resources and the equitable distribution of profits made from such research to local peoples. The protocol spells out how signatories

should take national legislative or policy measures to ensure that commercial work is done on 'mutually agreed terms' with local populations. To date, 41 governments have signed on, including Mexico and the European Union, but not the United States or China.

Although the protocol won't come into force until 90 days after the 50th nation signs up, countries such as Mexico are starting to explore how the agreement will play out. Souza is grateful for the treaty. "Without it, I could not guarantee that money would return to Cuatro Ciénegas on anything other than the good faith of researchers coming to the region," she says.

Among the more promising genes Souza has found so far are some that allow the breakdown of complex organic compounds, which could be useful for bioremediation, and some that enable the use of forms of phosphorus that are normally unavailable to life, which could be used to create plants that thrive without fertilizers. The latter, says Souza, is likely to result in the first patent from the region.

Groundwater is being extracted from Cuatro Ciénegas and from nearby valleys, mostly to grow alfalfa to feed dairy cows, lowering the local water table. That, combined with drought and natural variability, has dried out many ponds, says Elser, including one of the largest, which when full is more than a kilometre across and a metre deep. "It was actually gone this summer," Elser says. Souza estimates that her main study site won't last another 5 years at the current rate of water depletion.

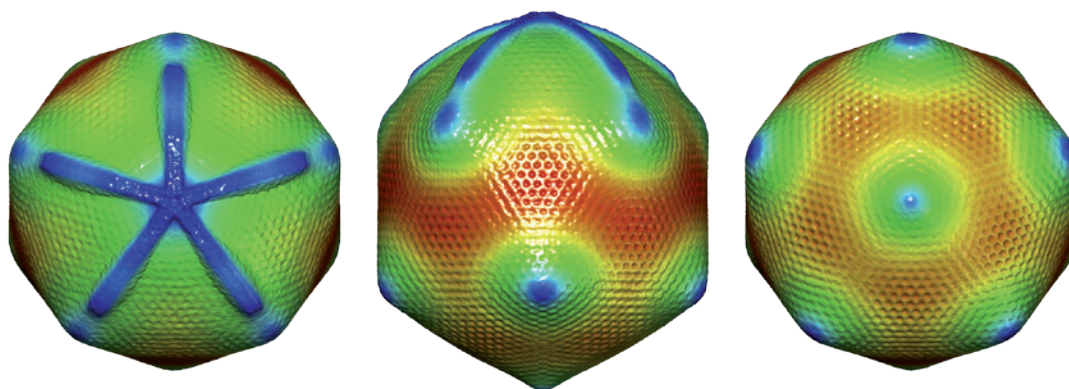
Over the past year, Souza has visited eight *ejidos* — communal land divisions — in the Cuatro Ciénegas area, explaining her intentions. Six have agreed to let her sample microbial life in their areas in exchange for a promise of future financial restitution; the other two wanted cash upfront, says Souza, which she is unwilling to provide. The percentage of profits that will be channelled back to communities is yet to be decided. Those communities will determine how funds are spent — Souza hopes on schools or on greenhouses that help to preserve water.

Getting a permit from the Mexican government wasn't easy, says Souza. Sonia Peña Moreno, a senior policy officer for biodiversity at the International Union for Conservation of Nature in Gland, Switzerland, who was involved with establishing the protocol, says that many researchers have hit brick walls when trying to navigate national laws on genetic resources. "That's very frustrating," she says, adding that many researchers are wary of additional bureaucratic complexity from the Nagoya Protocol.

It is unclear whether the vague promise of possible future funds will prove sufficient motivation to stop excessive groundwater extraction. Mexico's plans for wetland restoration are finally slowly getting under way, says Souza. In the meantime, she hopes her "biotechnology revolution for the people" will add to the cause. ■







# Out on a limb

*From giant viruses to unexplained marine DNA, there may be more to the tree of life than the three domains currently defined.*



The e-mail arrived two weeks after Jonathan Eisen's paper was published. Tongue-in-cheek, it read: "Welcome to the 'Fourth Domain' club." Eisen, an evolutionary biologist at the University of California, Davis, chuckled. His paper<sup>1</sup>,

which came out in March, hinted at bizarre new forms of microscopic life in the ocean. The e-mail was from Didier Raoult, a biologist at the University of the Mediterranean in Marseilles, France, who for years has argued that the colossal viruses he has identified belong to a unique, hitherto unknown, branch of life: the fourth domain.

Raoult believes that this domain branched off from the very base of life's evolutionary tree — a declaration that has been met with strong scepticism. Researchers disagree with the way he interprets sequence data. The debate highlights how hard it is to decipher the original relationships between microorganisms that have swapped, tinkered with and generally scrambled their DNA over billions of years.

Whether or not the fourth domain exists, the hunt for evidence has turned up several surprises. Raoult's microbes blur the boundaries between viruses — generally considered non-living things because they cannot function on their own — and the cellular organisms that host them, reigniting a debate about what constitutes life. Raoult says that some viruses are sophisticated enough to warrant special consideration. Eisen's approach, which sifted DNA from thousands of seawater samples, picked out sequences from — well, nobody quite knows what. "I would call it the dark matter of the biological universe," says Eisen of the countless unidentified DNA sequences lurking in the environment. "There is potentially enormous diversity out there, and this is the way to look for it."

## CELLULAR DIVISION

Before the advent of DNA sequencing, biologists divided life into two domains based on cellular structures. Eukaryotes, which encompass everything from amoebas to trees to humans, have large cells with internal structures such as nuclei and mitochondria. Prokaryotes are smaller and for the most part lack these structures. But in 1977, Carl Woese, a microbiologist at the University of Illinois at Urbana-Champaign, turned taxonomy upside down when he used RNA sequences to show

that prokaryotes comprise two very different groups: bacteria, which are distantly related to the eukaryotes, and archaea, which sit much closer to them on the evolutionary tree<sup>2</sup>. Woese faced

considerable backlash from the scientific community. It wasn't until the mid-1980s that his theory was accepted, bolstered by mounting genetic and molecular evidence. The textbooks were duly rewritten to feature three domains: eukaryotes, bacteria and archaea.

Raoult's claim for a fourth domain stems from his study of giant viruses. He and his colleagues identified the first of these<sup>3</sup>, Mimivirus, in 2003. The microbe had been found years before, infecting amoebas in the water of a cooling tower in Bradford, UK, but it was so big that it was mistaken for a bacterium. When Raoult and his colleagues couldn't identify it from its DNA, they looked at it under a microscope — and saw a virus.

The discovery stunned microbiologists. At 750 nanometres, Mimivirus was the first viral particle to be visible under a light microscope. "It was an exciting moment," says Jean-Michel Claverie, an evolutionary microbiologist at the University of the Mediterranean, who collaborated with Raoult on the discovery and analysis of Mimivirus. "People realized we didn't know anything about microbial biodiversity."

In 2004, Raoult and his group published the genome of Mimivirus<sup>4</sup>. At 1.2 million base pairs, its genome was at least twice as large as the largest known viral genome. And some of its more than 1,000 genes were involved in translating genetic information into proteins, something never before seen in viruses because they usually must rely on host translation machinery. The group drew up phylogenetic trees by comparing sequence information for seven proteins from Mimivirus with counterparts from organisms in the three domains. The giant virus, they wrote "appears to define a new branch", suggesting that it arose from an ancestor that could have predated the eukaryotes. According to Raoult, some viruses might deserve this special status on the tree because of their complexity and genetic make-up.

Objections ensued. David Moreira and Purificación López-García of the University of Paris-Sud in Orsay, France, included theirs in a list of ten reasons why viruses should be excluded from the tree of life<sup>5</sup>. For one thing, viruses can steal genetic material from their hosts in a

The outer capsid of Mimivirus: this giant was originally mistaken for a bacterium.

C. XIAO ET AL. PLOS BIOL. 7, E1000092 (2009)

process called horizontal gene transfer, which could be the source of the protein-translation genes.

Raoult, meanwhile, had found more giant viruses, including one in a Parisian cooling tower. Last year, he published a paper<sup>6</sup> that incorporated genetic data from the newly discovered giant viruses and from some other known large DNA viruses into phylogenetic trees. Because the sequences from these viruses seem to cluster together on their own branch of the tree, he argued, their genes must have been inherited together through evolution, and the group, called nucleocytoplasmic large DNA viruses (NCLDV), shared a common ancestor (see 'Family trees').

But resolving lineage using genetic sequences is not straightforward, particularly over great evolutionary distances. Over time, genes can mutate one way and later revert back to their former sequences, erasing the evolutionary trail of breadcrumbs that scientists try to follow. Two species can also independently evolve highly similar gene sequences, even though they aren't related, in a process called convergence. So organisms such as the NCLDVs could appear to belong together on close branches when really they are farther apart.

In June, Eva Heinz, an evolutionary biologist at Newcastle University, UK, and her colleagues published a response to Raoult's redrawn trees<sup>7</sup>. They reanalysed his data and built fresh trees based on models that account for artefacts such as convergence. The new trees showed the NCLDVs sprouting out as 'twigs' from several different branches, rather than being constrained to one branch. Heinz and her co-authors interpreted this to mean that the genes in the NCLDVs could have been stolen from various hosts on those branches through horizontal gene transfer. "With ancient phylogenetics, the model you use really matters," says Tom Williams, one of the co-authors, who is also at Newcastle. "When we used more biologically realistic models, we got a different tree that fitted the data better." And that tree had no fourth domain.

Raoult isn't convinced. It is true that large viruses do not branch together on the new trees, he says, but in his opinion, neither do the eukaryotes. "Maybe their analysis is not able to cluster organisms that have the same origin," he says. Heinz disagrees, saying that the eukaryotes do cluster together in their analysis. So Raoult is currently standing alone. "I'm ready to believe many exotic scenarios," says Claverie, "but my feeling is that at this point we don't have enough data, especially on those large viruses."

## GIANT HUNTING

Raoult is still trying to work out how his huge viruses evolved. One idea is that they could have come from a more complex ancestor that was later reduced to a parasitic shell. Raoult continues to search for new giant viruses and wants to use their genome sequences to work out when this group of viruses emerged, based on a molecular clock deduced from the viruses' mutation rates. And he avoids conversing with members of the virology community who are unwilling to challenge the 'dogma' of their field. "I'm more excited by discussions with the people trying to put a little bit of disorder in the way we think about biology," he says.

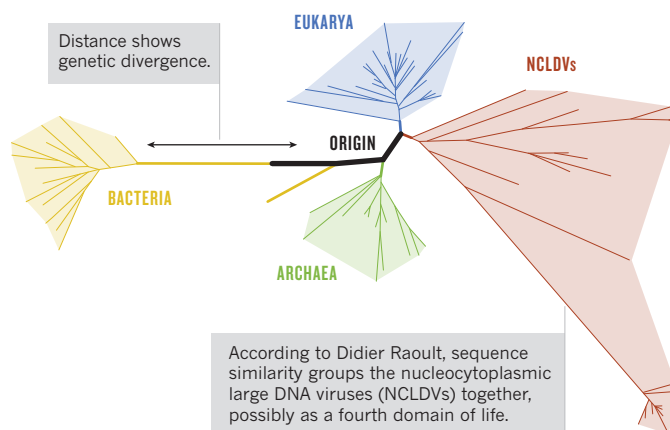
Woese says he can relate to this. "You've got to hang those ideas out there and take what comes," he says. But although Woese's iconoclastic ideas stood up to the criticism back in his day, he says he is less hopeful for Raoult's fourth domain.

So where does this leave Eisen? In his March study, he and his team probed seawater samples from the J. Craig Venter Institute's Global Ocean Sampling Expedition, an around-the-world venture that aims to find novel organisms by sequencing the mix of marine DNA. Most researchers doing environmental genomics search for organisms closely related to those already known, but Eisen cast a wider net. And when he drew up phylogenetic trees based on his haul, several unknowns stood out. A few looked like they might belong in Raoult's fourth branch, whereas others poked out from other branches. But with neither the time nor the funds to identify the

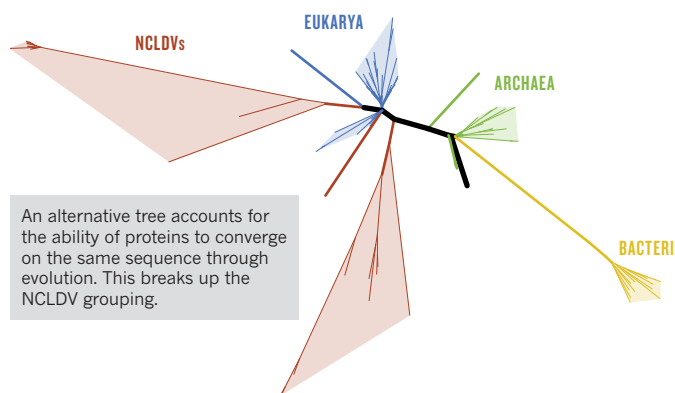
## Family trees

Phylogenetic trees try to show how species are related based on similarities in a common gene or protein sequence. In theory, the more similar the sequences, the more closely the organisms are related. But different models of the same sequence, such as the protein RNA polymerase II mapped here, can turn up different results.

### BRANCHING OUT



### A LIMB TORN ASUNDER



organisms containing the sequences, Eisen's team published the paper<sup>1</sup> detailing the methods and sequences in order to seek help from others.

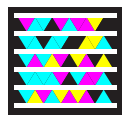
Eugene Koonin, a computational evolutionary biologist at the National Center for Biotechnology Information in Bethesda, Maryland, has done a preliminary analysis of the data that he says turned up nothing too unusual. "Quite a few of these sequences may represent very interesting new branches of bacteria and archaea," he says. But he ultimately sees it as a strong negative result: "They were hunting for the fourth domain but their net came back empty."

Eisen, who has yet to see Koonin's data, is not ready to sign up for Raoult's club. Rather, he says he published his data as a challenge to other researchers to be alert to what is truly unique in environmental samples, rather than discarding it as an artefact, as they often do. And there is plenty left to be discovered within the archaea and bacteria, says Francisco Rodríguez-Valera, a microbiologist at the Miguel Hernández University in Alicante, Spain. "There is a huge amount of microbial diversity that is unknown," he says. "I don't think we need to discover a new domain every ten years to convey to the general public the fact that microbes are important." ■

Gwyneth Dickey Zakaib writes for Nature from Washington DC.

1. Wu, D. *et al.* *PLoS ONE* **6**, e18011 (2011).
2. Woese, C. R. & Fox, G. E. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
3. La Scola, B. *et al.* *Science* **299**, 2033 (2003).
4. Raoult, D. *et al.* *Science* **306**, 1344–1350 (2004).
5. Moreira, D. & López-García, P. *Nature Rev. Microbiol.* **7**, 306–311 (2009).
6. Boyer, M. *et al.* *PLoS ONE* **5**, e15530 (2010).
7. Williams, T. A., Embley, T. M. & Heinz, E. *PLoS ONE* **6**, e21080 (2011).

➔ **NATURE.COM**  
Listen to the podcast  
for more on this story  
[go.nature.com/1nohq](http://go.nature.com/1nohq)



Scan the tag above with the  
free app from gettag.mobi





# GROWTH FACTOR

*Mu-ming Poo is nurturing a Shanghai neuroscience institute that offers a glimpse of his country's future as a bioscience superpower.*

BY DAVID CYRANOSKI

**M**u-ming Poo leads a double life. For three weeks every month, he works in a cramped, cluttered office at the University of California, Berkeley. Looking drab in his dark-green pullover, olive trousers and black Adidas sports shoes, the 62-year-old neuroscientist slumps slightly in his chair. In the adjoining laboratory, half a dozen post-doctoral researchers, expected to work independently, go quietly about their business.

Cut to Shanghai, China, where Poo spends the remaining quarter of his time. In the director's office at the Institute of Neurosciences (ION), he sports a pressed, light-blue shirt neatly tucked into belted trousers (same trainers). With few books and papers about, the room seems more spacious than its Californian counterpart; mangoes and other fruit in a bowl provide a tasteful flourish. Here, Poo supervises only one postdoctoral researcher, but a dozen chattering graduate students are stuffed into an office, waiting for the hour that he sets aside for each one during his whirlwind visits. Poo sits straighter, talks faster and seems more alert, alive — younger, even. As stimulating as he finds his research in the United States, where he is a member of the National

JON BERKELEY

Academy of Sciences, Poo finds a sense of mission in China. “It’s more exciting, exhilarating here,” he says. “They need me. I feel it’s the best use of my life.”

China is alive with possibilities in science, but realizing them is a complicated affair. The country’s fondness for speed — for short-term achievements and, increasingly, short-term profits — has worked relatively well in the chemical and physical sciences and in large-scale genomics, where researchers can systematically tick off the chemical compounds or genetic sequences that they have produced (see ‘Eastern promise’). But neuroscience and other fields of biology, which involve so many more unknowns, require more patience and a free-thinking academic culture that doesn’t mesh well with some entrenched features of Chinese academia. Graduate students in China typically become cogs in their supervisor’s machine, emulating his or her work. With quantity used as the yardstick for achievement, students churn out articles in minor journals. And principal-investigator positions are assigned by a bureaucratic tallying of young researchers’ publications. The best scientists, therefore, are almost guaranteed to go to the United States or elsewhere to progress.

But that is changing, thanks in no small part to Poo and the ION, a decade-old institute with 26 research groups, 350 members of staff and an outsize reputation. Poo is sometimes criticized as a micromanager and dictator (“Maybe a benevolent dictator,” he smiles), but ION students and group leaders revere him for his tireless mentoring, and for creating a US-style research culture geared towards top-level science. From his students, he demands critical thought and one first-author publication in a high-quality journal, rather than several in lesser ones. He evaluates faculty members and recruits on the basis of peer-reviewed publications and other achievements.

The results are clear. ION researchers authored China’s first neuroscience papers in *Cell*, *Science*, *Neuron*, *Nature Neuroscience* and *Nature Cell Biology* and, until 2005, the ION accounted for more than half of all top-level neuroscience publications from Chinese institutions, by Poo’s count. The ION’s “primary accomplishment has been to put neuroscience on the map in China in a way that no other basic neuroscience-research institute in the country has done”, says Richard Morris, a neuroscientist at the University of Edinburgh, UK, and a member of the institute’s international advisory board.

Over the past two years, the government and the Chinese Academy of Sciences (CAS) in Beijing awarded the institute three major funding programmes that will allow it to double in size by 2020. And the institute is pumping out some of the world’s sharpest, and most

aggressive, young neuroscientists, who are seeding its culture elsewhere. “It is exciting to see that high-quality neuroscience research has spread beyond the ION to other institutions throughout China,” says Bai Lu, who helped Poo to draft the proposal for the ION and is now vice-president of biology at the pharmaceutical company GlaxoSmithKline (GSK) China in Shanghai.

Even Westerners can see the tide turning. “It is only a matter of time before students will remain in China for their postdoctoral training,” says Thomas Insel, director of the US National Institute of Mental Health in Bethesda, Maryland. “And perhaps in a few years, we will be seeing US students and post-docs applying for positions in China.”

### THE DILETTANTE

Born on the mainland and raised in Taiwan, Poo earned a degree in physics at Taiwan’s elite Tsinghua University before moving to the United States, and into biophysics. His early studies of the movement of cell-membrane proteins led to an interest in events at the synapse, the junction between two neurons — and then onto synapse formation and plasticity, the process by which the junctions change with neuronal activity. His conversion from physicist to biologist defined

*“It’s more exciting, exhilarating here. They need me. I feel it’s the best use of my life.”*

thinks, an approach that he acknowledges has made him some enemies. And he is as direct when poking holes in the Chinese research system as when pinning down the next key experiment. He took that attitude to the ION, which opened in 1999. Although the institute was part of the crusty CAS, Poo’s mandate was to transform how science is practised in China by operating with autonomy and a Western style.

To replace the bureaucratic review of scientific work in China, for example, Poo brought in an independent review system that applied even to esteemed CAS members. “That was totally unheard of,” says Poo. One refused to be assessed, and left. In a review last year, two out of eight research groups at the ION got a ‘conditional’ pass, meaning that they need to publish within two years or risk dismissal.

The institute spans the range of basic neuroscience, from sensory mechanisms to neurodegeneration, stem cells and genetics. Following the US system, ION students — 20 or so graduate with PhDs each year — rotate through three different laboratories, evaluate their experiences and then choose a principal investigator as a thesis supervisor. Poo encourages students to critique articles in

a dilettante approach that he has maintained to this day. “I have a rather unconventional style,” says Poo. “I don’t work on one problem until it’s fully solved. I tend to skip around.”

Poo says what he

*Nature*, *Science* and *Cell*, which are often considered unassailable in China. He arranges for renowned scientists to give lectures, and asks his students to grill them with challenging questions. The ION was also the first CAS institute to allow a departing investigator to take, with some compensation, equipment and reagents to their next employer. “That offered the mobility that this country needs,” says Poo.

Meanwhile, Poo’s research continues full steam at Berkeley, where he has worked since 2000. He studies how neural activity shapes neural circuits, and increasingly how these mechanisms correlate with wider behavioural patterns such as learning and addiction.

In one project, Andrei Popescu, a postdoc in Poo’s Berkeley lab, is using mice to improve understanding of a phenomenon seen in children in the 1990s. Researchers at Rutgers University in Newark, New Jersey, found that when children were trained to overcome a hearing problem using rewards for success such as entertaining computer animations, they improved not just in the specific hearing skills, but also in broader language abilities<sup>1,2</sup>. Popescu and Poo hope to understand that process of ‘generalized’ learning by mimicking it in mice. Poo’s hypothesis is that learning a reward-associated task modifies the brain’s dopamine-based reward circuit, which is also thought to have a role in drug addiction<sup>3</sup>, in a way that makes an individual a better learner for other tasks. As Poo suggests, the subjects become “addicted to learning”.

Poo’s teaching philosophy mirrors his research hypothesis. He wants students to get hooked on research with the high of an early influential publication. “Those who get frustrated and give up — it’s because they didn’t have enough self-discipline to make the first success,” says Poo, who became hooked with his own paper in *Nature* in 1974 (ref. 4).

Poo works hard to shore up that discipline. Besides the monthly supervisory sessions with each of his ION graduate students, he holds weekly lab meetings over Skype. He reads all manuscripts from the ION to ensure that they are up to scratch. With two labs on different continents, not to mention an entire institute to run, how does he fit it all in? “I put in more time, more effort,” he says. “I take no weekends. I work like a dog. I’m here every night.”

### THE TAO OF POO

Poo’s reputation took a public beating in 2002, when one of his Berkeley lab members posted online an e-mail in which Poo lamented the group’s poor progress. “If there is no drastic change in the lab, Poo lab will soon cease to be a productive, first-rate lab that you chose to join in the first place,” he wrote, before imposing strict rules, including a minimum 50-hour, 6-day working week spent mostly at the bench (reading papers was for after-hours). Anyone who chose not to follow the rules should “start making plans immediately and leave the lab”.

**➔ NATURE.COM**  
To read more about the rise of Chinese science, visit:  
[go.nature.com/kumot7](http://go.nature.com/kumot7)



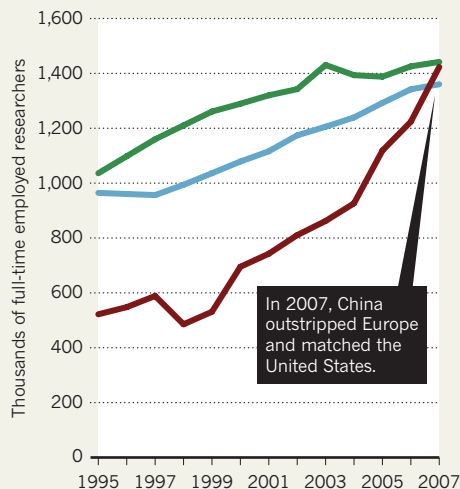
## EASTERN PROMISE

Chinese science is flourishing, but still has a way to go.

— United States — European Union — China

## POPULATION BOOM

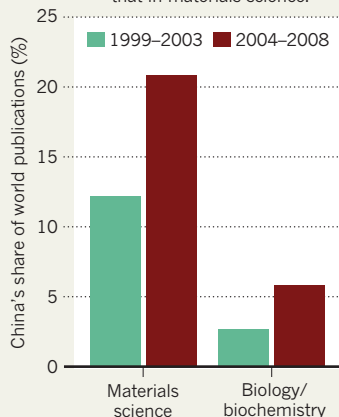
The number of scientists has grown faster in China than elsewhere.



In 2007, China outstripped Europe and matched the United States.

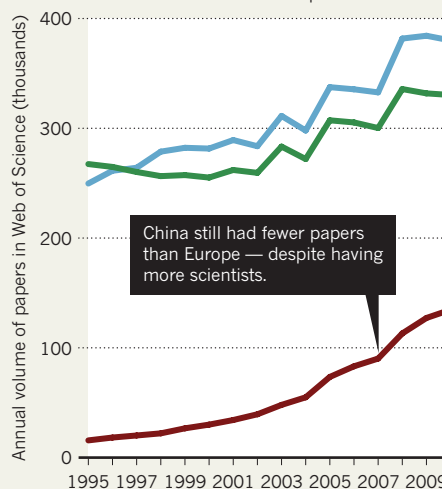
## MATERIAL WORLD

Progress in biology lags behind that in materials science.



## PAPER TRAIL

Scientific publications have increased around the world in the past decade.



China still had fewer papers than Europe — despite having more scientists.

One student did.

Poo stands by his words. “Young people, if they want to make it in sciences, must work hard,” he says. Did the letter have an effect on his ability to get graduate students? Poo pauses. “It had a screening effect.” His last Berkeley PhD student graduated last year, and he hasn’t actively recruited since. “I don’t have the time to mentor them,” he says. The lab has shrunk from some 20 people in 2007 to just 6.

Popescu says that images of Poo as a dictator are off-base. Two conversations with former lab members quickly reassured him before he joined. “I thought it better to trust people who had been there than something posted on the Internet,” he says. Poo’s dozen Shanghai students, who come from nearly as many provinces around China, describe an intense but supportive and tolerant atmosphere. One boasts about starting work days at noon (although admits working six or seven days per week). Hailan Hu, a principal investigator at the ION, says that Poo’s meticulous attention — he “doesn’t just fix the paper but also teaches grammar and presentation skills” — is “difficult for him but great for the students. You need to book him in advance”.

Insel recalls a visit to the ION a couple of years ago that “left quite a lasting impression”. Challenging questions from graduate students and the intense focus of the faculty members, he says, “all left me with a very profound sense of respect for what Mu-ming Poo and his colleagues were building in Shanghai”.

## FERTILE GROUND

In 2009, the CAS approved a ten-year plan to expand the ION to 50 laboratories. In the same year, it selected the centre as a ‘trial institute’ for stable funding and doubled the support per investigator from 600,000 renminbi

(US\$93,000) to 1.2 million renminbi. Stable funding now covers 40–50% of the ION budget, in comparison with 30% before. Research grants make up most of the rest. Poo scored an even bigger coup this year, when his ‘neural basis of intelligence’ project was selected to receive one of China’s largest and most sought-after basic-science grants, as part of the science ministry’s National Basic Research Program. Only 11 projects were selected this year, and just 3 in biological sciences. With 80 million renminbi for the next five years and a promise that the project will continue for another ten, Poo is now building a national programme and recruiting

*“Young people, if they want to make it in sciences, must work hard.”*

ing a dozen research groups from ten institutions, in addition to the existing ION groups. Signs of China’s impending boom in neuroscience are evident elsewhere. In 2008, East China Normal University in Shanghai launched an Institute of Cognitive Neuroscience, full of returnees from the United States, and started planning a large primate centre. This April, Patrick McGovern, a technology and media tycoon who bankrolled the McGovern Institute for Brain Research at the Massachusetts Institute of Technology in Cambridge, agreed to donate \$10 million to form a sibling institute at Tsinghua University in Beijing. He is negotiating similar agreements with Beijing Normal University and Peking University.

But the rapid progress in Chinese neuroscience is creating problems for the institute that is driving it. GSK’s international neurodegenerative-disease research base — attracted to Shanghai partly because of the ION — has recruited at least seven graduate students from

the institute, many of whom, Poo laments, left without finishing a doctorate. The ION has also lost senior staff to the National Institute of Biological Sciences in Beijing, and other universities. Poo blames the departures on better salaries elsewhere, although a critic — who does not want to be named — points to his ‘overly controlling’ managerial style. Either way, more than half of the researchers at the ION have been there for fewer than five years; and publications in leading journals have dropped. Poo created a fluid system, and the people flowed. “But I don’t worry,” he says. “It’s a place to start your career. It’s a nursery.”

Poo is still sceptical about the future of Chinese science. He worries that misconduct is still tolerated and that the country’s work ethic is being eroded, with students demanding comfortable living arrangements, better food and vacations. And Poo pooh-poohs the government’s readiness to give huge packages to scientists solely because they have worked overseas. “That’s no way to cultivate native talent,” he says.

Listen to Poo talk about China, and it sounds as if the country might not soon become a global force in biological sciences. But watch what he does, and the signs of the country’s emergence in this field are clear. “I have 40 years of experience in first-rate institutions,” he says. “If you want this to be a first-rate institution, follow me.”

And China is following. ■ SEE EDITORIAL P.5

David Cyranoski is Nature’s Asia-Pacific correspondent.

1. Merzenich, M. M. *et al. Science* **271**, 77–81 (1996).
2. Tallal, P. *et al. Science* **271**, 81–84 (1996).
3. Liu, Q.-S., Pu, L. & Poo, M.-M. *Nature* **437**, 1027–1031 (2005).
4. Poo, M.-M. & Cone, R. A. *Nature* **247**, 438–441 (1974).

# COMMENT

**BIOETHICS** Growth in genome screening could cause dangerous meddling **p.27**

**EVOLUTION** How genes and culture have shaped our ability to cooperate **p.29**

**CHEMISTRY** Debating how life got going on the early Earth **p.30**



**EXHIBITION** Wildlife paintings from Yukon to Yellowstone **p.32**



ILLUSTRATION BY JONATHAN BURTON

## Search needs a shake-up

On the twentieth anniversary of the World Wide Web's public release, **Oren Etzioni** calls on researchers to think outside the keyword box and improve Internet trawling.

Two decades after Internet pioneer Tim Berners Lee introduced his World Wide Web project to the world using the alt.hypertext newsgroup, web search is on the cusp of a profound change — from simple document retrieval to question answering. Instead of poring over long lists of documents that contain requested keywords, users need direct answers to their questions. With sufficient scientific and financial investment, we could soon view today's keyword searching with the same nostalgia and amusement reserved

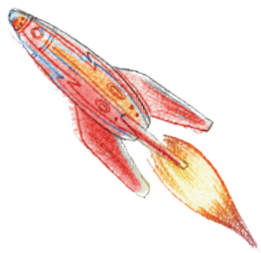
for bygone technologies such as electric typewriters and vinyl records.

But this transformation could be unreasonably delayed. As a community, computer scientists have underinvested in tools that can synthesize sophisticated answers to questions, and have instead focused on incremental progress in lowest-common-denominator search. The classic keyword search box exerts a powerful gravitational pull. Academics and industry researchers need to achieve the intellectual 'escape velocity' necessary to revolutionize

search. They must invest much more in bold strategies that can achieve natural-language searching and answering, rather than providing the electronic equivalent of the index at the back of a reference book.

Today, that 'book' is distributed over billions of web pages of uneven quality, and much effort has been directed at ranking the most useful results. Such engines readily index billions of documents, but overwhelm their users with millions of results in response to simple queries. This quandary only worsens as the number of web pages ►





► grows, and as web access shifts to mobile devices with tiny screens.

Moving up the information food chain requires a search engine that can interpret a user's question, extract facts from all the information on the web, and select an appropriate answer.

The big search engines have taken tiny steps in the right direction. Google directly displays film showing times and weather forecasts in response to particular queries, but this is a drop in the ocean of possible search queries. Bing markets itself as a 'decision engine' rather than a 'search engine', but with the exception of its airfare predictor, its differences from Google are limited. Wolfram Alpha's 'computational knowledge engine' provides remarkably sophisticated answers to certain questions. For example, when asked "am I too drunk to drive?", Wolfram Alpha offers to compute your blood alcohol level based on factors such as your weight and the number of drinks consumed. Unfortunately, this approach applies to only a very limited set of pre-specified questions (it fails to answer the similar question "have I had too much to drink?").

Microsoft acquired Powerset, a start-up company developing a natural-language search engine, for upwards of US\$100 million in 2008. And in 2010, Apple bought Siri, a start-up with an iPhone app that answered natural-language questions about films, restaurants and other local services, for a price speculated to have been more than \$200 million. But these efforts are dwarfed by companies' investments in keyword search. In short, search engines have yet to develop general-purpose question-answering capabilities.

### INFORMATION EXTRACTION

Work on the challenge of automatically recovering factual information from text began in the 1970s. Early information-extraction systems were hand-crafted to particular genres and very narrow topics. For example, the JASPER system extracted financial information from Reuters news-wire text in the 1980s. Broadening their scope was both labour-intensive and error prone. In the 1990s, a more automated class of information-extraction system emerged. Instead of relying on hand-crafted rules to glean facts from sentences, the systems generated the rules automatically based on a curated collection of example

sentences. This automated approach is more streamlined and less error prone, but still requires careful manual effort to create an example collection for each topic of interest.

In 2007, my lab introduced open information extraction — methods that scale to any topic and to arbitrary English sentences. The basic idea is remarkably simple: most sentences contain highly reliable syntactic clues to their meaning. For example, relationships are often expressed through verbs (such as invented, married or elected) or verbs followed by prepositions (such as invented by, married to or elected in). It is often quite straightforward for a computer to locate the verbs in a sentence, identify entities related by the verb, and use these to create statements of fact. Of course this doesn't always go perfectly. Such a system might infer, for example, that 'Kentucky Fried Chicken' means that the state of Kentucky fried some chicken. But massive bodies of text such as the corpus of web pages are highly redundant: many assertions are expressed multiple times in different ways. When a system extracts the same assertion many times from distinct, independently authored sentences, the chance that the inferred meaning is sound goes up exponentially.

Open information extraction obviates topic-specific collections of example sentences, and instead relies on its general model of how information is expressed in English sentences to cover the broad, and unanticipated, universe of topics on the Internet.

Other approaches to information extraction are also yielding important results. There are projects at Carnegie Mellon University in Pittsburgh, Pennsylvania, Stanford University in Palo Alto, California, and New York University, at companies such as Google and Microsoft, and at numerous start-ups. In contrast to open information extraction, however, these methods cannot automatically operate at the scale of the web. For instance, some projects have come up with ultrafast systems to provide in-depth understanding of sentences, but they only work in specific domains such as finance, or in particular genres such as Wikipedia articles.

Some scientists are experimenting with 'power tools' that delve into the content of scientific articles to suggest novel connections and potential hypotheses (see *Nature* **463**, 416–418; 2010). Many of these tools are not fully automated, however, which immediately leads to challenges in extending them beyond carefully circumscribed arenas such as gene names in PubMed abstracts. The

use of open information extraction would substantially broaden their scope. The open-source code for our system is available at [go.nature.com/ei3p4f](http://go.nature.com/ei3p4f).

Much more research has to be done to improve information-extraction systems — including our own. Their abilities need to be extended from being able to infer relations expressed by verbs to those expressed by nouns and adjectives. Information is often qualified by its source, intent and the context of previous sentences. The systems need to be able to detect those, and other, subtleties. Finally, automated methods have to be mapped to a broad set of languages, many of which pose their own idiosyncratic challenges.

### PROGRESS IN JEOPARDY

The main obstacle to the paradigm shift from information retrieval to question answering seems to be a curious lack of ambition and imagination. Much of the research on natural language processing is focused on limited tasks, such as recovering the syntactic structure of sentences rather than trying to uncover their meaning, or on methods that do not scale to massive corpora and arbitrary topics because of their reliance on manually-annotated data, or on algorithms whose computation grows explosively with the amount of text involved.

In 2009, the US Defense Advanced Research Projects Agency began a 'Machine Reading' programme that has focused attention on this area, with tens of millions of dollars awarded in contracts to a handful of teams — including mine. An order of magnitude more research funding is necessary, as is a focus on scaling up current methods to the size and heterogeneity of the web.

One exceptional system — IBM's Watson — utilizes a combination of information extracted from a corpus of text equivalent to more than 1 million books combined with databases of facts and massive computational power. Watson won a televised game of Jeopardy against two world-class human players in February this year. The multi-billion dollar question that IBM is now investigating is 'can Watson be generalized beyond the game of Jeopardy?'

General-purpose question-answering systems will be a boon to scientists searching the literature, and to the increasing number of us who access the web's richness through a mobile phone with a tiny screen that necessitates concise responses. Without it, we risk drowning in the growing sea of information. ■

**Oren Etzioni** is the director of the Turing Center at the University of Washington, Seattle, Washington 98195, USA.  
e-mail: [etzioni@cs.washington.edu](mailto:etzioni@cs.washington.edu)

*"The main obstacle to question answering seems to be a curious lack of ambition and imagination."*



# Growth of genome screening needs debate

There could be unexpected consequences if greater understanding of disease genetics gives parents more choice in what they pass to their children, says **David B. Goldstein**.

**T**wenty years ago, at the outset of the human genome project, tens of genes were known to cause Mendelian diseases when disrupted. These rare conditions are usually caused by mutations in single genes, and are passed on from one or both parents according to simple inheritance laws. Now, nearly 3,000 of these genes have been mapped. Advances in screening mean that prospective parents could in principle find out whether they are carriers of each disease-causing mutation in any of these genes before choosing a reproductive partner or before conceiving. Couples could also find out whether an embryo created through *in vitro* fertilization carries such mutations, or whether a fetus does. Indeed, a fetus's entire DNA sequence can now be determined by sequencing the fragments of fetal DNA in the mother's blood<sup>1,2</sup>.

Screening will not stop here. Until recently, the idea that you could identify very strong

genetic predictors of a person's risk of developing common diseases — such as epilepsy or type 2 diabetes — through screening seemed unlikely because of the presumed complexity of the underlying genetics. Evidence now suggests that numerous genetic variants, each rare in the general population, have a strong influence on which common diseases people get and when. Over the next five years, the sequencing of hundreds of thousands of human genomes is likely to identify many gene variants that, individually or in combination with a few other variants, have a strong effect on disease predisposition.

Societies and governments are ill-prepared for the imminent availability of this information. We must consider now the extent to which we want to dictate the genomes of our children, the individual and societal implications of doing so, and where the authority for such decisions should rest — with governments or parents?

Following the Second World War, a consensus emerged among biologists that the genetic bases of most common diseases are fundamentally different from those of Mendelian ones. Many thought that these diseases, instead of being caused by severe rare disruptions to genes, were strongly influenced by the combined effect of numerous common variants — mutations found in more than, say, 1% or 5% of the population. Each of these mutations was thought to have either small effects on a carrier's risk of developing a disease or nuanced effects — conferring problems in some environments and perhaps benefits in others.

There was comfort in complexity in this world view. Although screening might be appropriate to curb rare Mendelian conditions such as Tay-Sachs disease and cystic fibrosis, such an approach seemed inconceivable for common 'complex' diseases.

In the past few years, support for this

ILLUSTRATION BY DAVID PARKINS



picture has waned. Genome-wide association studies interrogate a set of known variants, most of which are strongly associated with others. By investigating the standardized set, geneticists can obtain information about virtually all the common variants in the human genome. Some studies have scanned tens of thousands and, recently, hundreds of thousands of genomes. Yet for most diseases, only fractions of the genetic control have been explained. What's more, in cases where geneticists have identified many variants, each having a minuscule effect, it is difficult to determine which ones have a causal role.

An even stronger challenge to the idea that numerous common variants underlie common diseases comes from the study of copy-number variants. These deleted or duplicated stretches of DNA sometimes involve scores of genes. They are far more similar than expected to Mendelian-disease mutations — both in being obviously harmful, and in having large effects on a carrier's risk of developing a disease. A small but rapidly growing number of very rare copy-number variants have been identified as definitive risk factors for several common diseases, including autism and epilepsy<sup>3,4</sup>.

In other words, at least some common diseases may turn out to be caused by numerous rare genetic mutations, each with considerable effect, that differ from person to person.

## HARD TO TREAT

Many in the medical-genetics community stand by the idea that common variants are the primary driver of common disease<sup>5,6</sup>. Yet the perspective I've outlined here — that some common diseases are strongly influenced by rare high-impact mutations — has gained considerable support over the past four years<sup>7</sup>.

To me, the most troubling implication of this is the inevitable expansion of screening that will happen once these rare mutations are discovered.

One of the most sobering lessons from the study of Mendelian diseases is that identifying underlying genetic defects rarely leads to the rapid development of effective treatments. Success stories exist: the painful bone lesions and anaemia of enzyme deficiencies such as Gaucher's disease, for example, can be greatly reduced by injecting patients with artificially synthesized enzyme. But many more diseases remain incurable even though their genetic bases have long been known. In some cases, such as cystic fibrosis, care regimes have relieved symptoms and extended lifespan, but these advances are not cures and so far have not been much informed by knowledge of the responsible gene.

Finding treatments for non-Mendelian diseases could prove just as difficult. In the meantime, the demand for screening to give prospective parents the option of not transmitting mutations is likely to

soar — especially given the plummeting costs of sequencing entire genomes. (The US\$4,000–\$5,000 needed today roughly matches the cost of some widely used medical imaging procedures.)

Currently, there is no consensus on which mutations should be identified through screening. Tests are routinely offered in the United States and Europe for mutations causing early onset, serious Mendelian diseases such as Tay Sachs, Down's syndrome and cystic fibrosis. Indeed, since the late 1970s, screening — carried out either before conception, before an embryo is implanted or during pregnancy — has reduced the number of children born with Tay Sachs in the United States by 90%. Even screens for later-onset conditions such as Huntington's disease are offered in some US fertility centres. Yet no screens are available for the *APOE4* variant that confers a

**"Many diseases are incurable even though their genetic bases have long been known."**

dramatically increased risk of late-onset Alzheimer's disease.

Testing for variants that may not necessarily lead to disease, or that underlie diseases of adulthood or old age, may seem less urgent than testing for ones that mean a child will die young. But some of the people who carry the *APOE4* allele today would like to be sure they don't transmit it.

Today, decisions about what should be tested for tend to be made by government organizations, such as the UK Human Fertilisation and Embryology Authority, or by medical practitioners in fertility clinics, as in the United States. The German parliament is considering allowing prenatal genetic testing in only a few situations, for instance when pregnancy would probably result in stillbirth or miscarriage. I am supportive of the rights of parents to choose whether they wish to transmit variants such as *APOE4* that are strongly associated with risk of disease, although this inevitably raises the question of where to draw the line.

## TOUGH CHOICES

Within the next few years, our ability to identify pathogenic and potentially pathogenic mutations — as well as the huge number of mutations that no one can vouch for as dangerous or safe — will almost certainly outstrip our ability to act on the information. For example, if parents using *in vitro* fertilization wanted to avoid transmitting five mutations, physicians would have to screen scores of embryos to have a reasonable chance of finding one that carried none of the mutations. Likewise, parents would have to terminate an unfeasible number of pregnancies to be assured that their fetus was unaffected.

Such constraints are unlikely to apply indefinitely, however. It is already routine

for researchers to selectively 'edit' DNA sequences in certain kinds of cell. This cannot yet be done in human sperm and eggs, but several technological advances that are generally expected would permit the effective editing of gamete DNA<sup>8</sup>. This would allow a qualitatively different kind of screening.

Rare variants are more likely to damage genes than common ones<sup>9</sup>. Data generated from various whole-genome sequencing studies carried out at the Center for Human Genome Variation at Duke University in Durham, North Carolina, suggest that every person carries hundreds of distinct protein-changing variants that exist in less than 1% of the general population. Might some parents choose not to transmit any such rare variants to 'play it safe' if they could?

One potential problem with this is that numerous genetic risk factors will have diverse and unexpected effects — sometimes causing disease, sometimes being harmless and sometimes perhaps being associated with behaviours or characteristics that society deems positive. Even for simpler Mendelian diseases, up to 30% of the mutations originally termed pathogenic have turned out to be apparently harmless<sup>10</sup>. Wholesale elimination of variants associated with disease could end up influencing unexpected traits — increasing the vulnerability of populations to infectious diseases, for instance, or depleting people's creativity.

There are no clear-cut answers to the questions of what should be screened for and to what end, but we must at least begin the debate. ■

**David B. Goldstein** directs the Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA.  
e-mail: d.goldstein@duke.edu

- Lo, Y. M. et al. *Sci. Transl. Med.* **2**, 61ra91 (2010).
- Fan, H. C. & Quake, S. R. *PLoS ONE* **5**, e10439 (2010).
- Merikangas, A. K., Corvin, A. P. & Gallagher, L. *Trends Genet.* **25**, 536–544 (2009).
- Heinzen, E. L. et al. *Am. J. Hum. Genet.* **86**, 707–718 (2010).
- Lander, E. S. *Nature* **470**, 187–197 (2011).
- International Schizophrenia Consortium *Nature* **460**, 748–752 (2009).
- McClellan, J. & King, M. C. *Cell* **141**, 210–217 (2010).
- Potter, S. *Designer Genes: A New Era in the Evolution of Man* (Random House, 2010).
- Zhu, Q. et al. *Am. J. Hum. Genet.* **88**, 458–468 (2011).
- Bell, C. J. et al. *Sci. Transl. Med.* **3**, 65ra4 (2011).

## CORRECTION

The Comment article 'The unplanned impact of mathematics' (*Nature* **475**, 166–169; 2011) credited Andrew Odlyzko with publishing results on kissing numbers in 8 and 24 dimensions. This work was done jointly with Neil Sloane.



A. RANTE/BARCROFT MEDIA

Showing high levels of cooperation, whale hunters from Lamalera in Indonesia work together to spear a whale.

## EVOLUTION

# Not so selfish

A prescription for how human cooperation evolved will provoke much-needed debate about the origins of society, finds **Peter Richerson**.

**H**umans are capable of remarkable feats of cooperation. Warfare is an extreme example: when under attack, hundreds or even millions of people might join forces to provide a mutual defence. In *A Cooperative Species*, economists Samuel Bowles and Herbert Gintis update their ideas on the evolutionary origins of altruism. Containing new data and analysis, their book is a sustained and detailed argument for how genes and culture have together shaped our ability to cooperate.

Modern hunting and gathering societies offer clues as to how human cooperation evolved. They are typically organized into tribes of a few hundred to a few thousand people. Each tribe is composed of smaller bands of around 75 individuals united by bonds of kinship and friendship. Formalized leadership is often weak, but cooperation is buttressed by social norms and institutions, such as marriage, kinship and property rights. The tribal scale of social organization probably evolved by the late Pleistocene (126,000–11,700 years ago), or perhaps much earlier.

Human societies are diverse and competitive, often violently so. Charles Darwin conjectured in *The Descent of Man* (John Murray, 1871) that the main evolutionary motor behind human cooperation was intertribal competition, and suggested that cooperation evolved in two stages. In 'primeval' times, well before the dawn of recorded history, our ancestors came under selection for cooperative instincts, such as sympathy and group loyalty. In more recent 'civilized' times, laws and customs have fostered cooperation

on ever larger scales. Darwin contended that the primeval social emotions, more than natural selection, drove the evolution of civilization.

Around 1900, concern with the Darwinian mechanisms of evolution died out in the emerging social sciences, shortly before the topic exploded in biology. Sociobiologist E. O. Wilson and others resurrected interest in the 1970s. Around the same time, geneticists Luigi Luca Cavalli-Sforza and Marcus Feldman introduced the methods of population genetics to the study of cultural evolution.

Bowles and Gintis seek to explain, as did Darwin, why humans have moral sentiments. Most people care about the welfare of others, even beyond kin and close associates. Economists use experimental games to test the assumption that humans are self-interested actors, and often find it wanting. Many participants show fairness and cooperation, punishing those who behave selfishly. The term 'strong reciprocity' is used to describe the tendency of people to sacrifice their own payoffs by rewarding or punishing others. Even in anonymous groups in the laboratory, strong reciprocity



### **A Cooperative Species: Human Reciprocity and Its Evolution**

SAMUEL BOWLES AND HERBERT GINTIS  
Princeton University Press: 2011. 288 pp.  
\$35, £24.95

deters free-riding and sustains cooperation.

Bowles, Gintis and their colleagues have used games to probe the moral sentiments of a range of societies in Africa, Indonesia and South America. They find that cultures vary greatly in their social preferences. Nowhere do individuals, on average, act with complete self-interest, but how much they will pay to punish others varies. Moreover, the groups' experimental behaviour correlates with that in the real world. Societies living in hamlets within tropical forests who cooperate little in real life also show little regard for the welfare of others in the games. By contrast, an Indonesian society of whale hunters rates highest on strong reciprocity.

The lack of completely selfish societies and the variability in strong reciprocity across different societies implies that the evolution of cooperation involves both genes and culture. The question is how the two systems interacted during our evolutionary history.

*A Cooperative Species* attacks this problem theoretically and empirically. After introducing the mathematical models that have been used to explain the evolution of social behaviour in humans and other animals, the authors caution that some are implausible. For example, game theorists have discovered that these games have many solutions and different groups might choose different rules to govern and reward behaviour by negotiation among self-interested parties. Bowles and Gintis point out that the solution adopted by a group emerges as a result of historical processes and is hard to change by negotiation or persuasion. ▶



► To illustrate how cooperative behaviour may have arisen, the authors build simple computer simulations of Pleistocene human societies. Plausibly, they propose that the crucial step towards human social systems was the evolution of a cooperative unit that was big enough to insure against the risks involved in hunting large game, comprising around 32 adults plus juveniles and the elderly. Such bands, the authors argue, would have a modest amount of genetic variation between them. So, even if violent conflict between bands was common, group selection could not favour costly altruistic acts.

Selection among groups for cooperation gains traction only where it exceeds that for selfishness within a group. Social institutions can reduce the advantages of selfish behaviour. For example, modern hunter-gatherers typically share their resources. More successful providers are 'taxed' in food to support the collective. This sharing limits variation in reproductive success within groups, imposing selection on groups for individually costly acts of cooperation.

The authors' modelling shows that other institutions — such as ostracism, coordinating punishment of defectors and restricting altruism to the local genetically related group — can have the same effects. They suggest that selection can favour emotions such as shame or guilt that internalize social norms and benefit the group but cost the individual.

Although *A Cooperative Species* is broadly representative of the gene-culture co-evolutionary approach to human cooperation, I beg to differ on some points. In my view, the critical late-Pleistocene groups in which altruism should be explained are the larger tribes composed of many bands. These have the crucial feature of substantial cooperation between genetically unrelated individuals, on which the evolution of complex societies is based. Although bands do sometimes have violent conflicts, intratribal relations are usually more peaceful than intertribal ones. In my opinion, the authors also accept too high a value for the genetic differences between neighbouring populations in their simulations. However, such a discussion illustrates the book's strength. By presenting clear models that are tied tightly to empirically derived parameters, Bowles and Gintis encourage much-needed debate on the origins of human cooperation. ■

**Peter Richerson** is professor emeritus in the Department of Environmental Science and Policy, University of California, Davis, California 95616, USA.  
e-mail: [pjricherson@ucdavis.edu](mailto:pjricherson@ucdavis.edu)



T. HOFFMAN

Biochemist David Deamer poured chemicals into a hot pool in 2005 to see if primitive cells would form.

## ASTROBIOLOGY

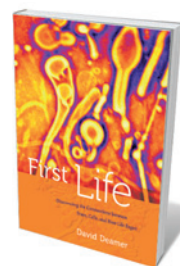
## Life's beginnings

Robert Shapiro on a reminder that laboratory experiments don't always translate to nature.

In June 2005, an group of international scientists clustered around a small, near-boiling pool in a volcanic region of Siberia. Biochemist David Deamer took a sample of the waters, then added to the pool a concoction of organic compounds that probably existed 4 billion years ago on the early Earth. One was a fatty acid, a component of soap, which his laboratory studies suggested had a significant role in the origin of life.

Over several days, Deamer took many more samples. He wished to see whether the chemical assembly process that he had observed in his laboratory, which eventually produced complex 'protocell' structures, could also take place in a natural setting. The answer was a resounding no. The clays and metal ions present in the Siberian pool blocked the chemical interactions.

This experiment was a reality check, explains Deamer in *First Life*. He proposes in the book that the complex molecules that led to life developed not in 'warm little ponds', but in tiny droplets bound by fatty acids. Although his account lacks the tales of personality and conflict that enliven other works in astrobiology, Deamer's



**First Life:**  
Discovering the  
Connections  
between Stars,  
Cells, and How Life  
Began

DAVID DEAMER  
University of California  
Press: 2011. 288 pp.  
\$28.95, £19.95

demonstration that we cannot translate lab results to natural settings is valuable.

Because we can get reactions to work in the controlled conditions of a laboratory, he cautions, it does not follow that similar ones occurred on prebiotic Earth. We might overlook something that becomes apparent when we try to reproduce the reactions in a natural setting. This provocative insight explains why

the origin-of-life field has been short on progress over the past half century, whereas molecular biology has flourished.

Today, the simplest living cells depend on molecules that are far more intricate than those that have been isolated from sources unrelated to life (abiotic), such as meteorites. The most noteworthy chemical substances in life are functioning polymers — large molecules made of smaller units called monomers, connected in a specific order. The nucleic acids RNA and DNA, carriers of

➔ **NATURE.COM**  
For more on the  
RNA-world  
hypothesis, see:  
[go.nature.com/kuolyb](http://go.nature.com/kuolyb)

genetic information and heredity, are made of connected nucleotide monomers. Similarly, proteins are vital polymer catalysts that are made by combining monomer amino acids. Such modern biological constructions were unlikely to have been present on the early Earth.

Despite this, many researchers have tried to demonstrate that RNA, or something similar, turned up spontaneously between 3 billion and 4 billion years ago. Physicist and biochemist Walter Gilbert suggested in 1986 that life began with the spontaneous generation of an RNA that could copy itself: the 'RNA world'. The advantage of this idea is

**"The advantage of the 'RNA world' idea is that one polymer would be all that was needed to get life started."**

that the formation of just one polymer would be all that was needed to get life started. The disadvantage is that such an event would be staggeringly improbable.

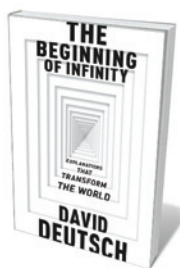
Nucleotides, for example, are not encountered in nature beyond organisms or laboratory synthesis. To construct RNA, high concentrations of four select nucleotides would be needed in the same location, with others being excluded. If this is the prerequisite for life, then it is an unusual phenomenon, rare in the Universe. As an alternative, other scientists (myself included) have suggested that life started without the presence of polymers; that instead, heredity and catalysis began with monomers.

Deamer's thesis diverges from the standard RNA-world concept. He focuses not on the generation of a naked RNA-like polymer, but on the formation of a simple cell-like compartment, or vesicle. Modern cells are enclosed by a complex fatty membrane, which prevents leakage. Vesicles with similar properties have been formed in the lab from certain fatty acids. Deamer holds that the spontaneous formation of vesicles, into which RNA could be incorporated, was a crucial step in life's origin. Unfortunately, his theory retains the improbable generation of self-replicating polymers such as RNA.

Nevertheless, Deamer's insight deflates the synthetic proofs put forward in numerous papers supporting the RNA world. He ends *First Life* by calling for the construction of a new set of biochemical simulators that match more closely the conditions on the early Earth. Unfortunately, the chemicals that he suggests for inclusion are drawn from modern biology, not from ancient geochemistry. We should let nature inform us, rather than pasting our ideas onto her. ■

**Robert Shapiro** was professor emeritus of chemistry at New York University. Sadly he passed away on 15 June 2011, shortly after completing this review.

## Books in brief



### The Beginning of Infinity: Explanations That Transform the World

David Deutsch ALLEN LANE 496 pp. £25 (2011)

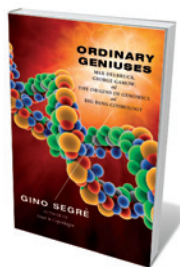
Scientific explanations have an infinite scope and are ever-adaptable, argues quantum computation expert David Deutsch in his latest book, which is sure to provoke many philosophers of science. Since the Enlightenment of the eighteenth century, the scientific method has allowed us to continually describe, assess and reconfigure ideas about the Universe, in a virtuous cycle that Deutsch sees as boundless. Everything is within the reach of reason, he claims, from free will to creativity and the laws of nature. It is our duty to seek the best explanations, he says.



### The Sorcerers and Their Apprentices: How the Digital Magicians of the MIT Media Lab Are Creating the Innovative Technologies That Will Transform Our Lives

Frank Moss CROWN 272 pp. \$27.50 (2011)

Frank Moss, a former director of the Media Lab at the Massachusetts Institute of Technology (MIT) in Cambridge, reflects on five years at the helm of this innovative institution. Through tales of the people who worked on such imaginative projects as the development of child-safe air bags and Lego robots, he highlights the Media Lab's ethos of creative freedom, serendipitous discovery and porous disciplinary boundaries.



### Ordinary Geniuses: Max Delbrück, George Gamow, and the Origins of Genomics and Big Bang Cosmology

Gino Segrè VIKING 352 pp. \$27.95 (2011)

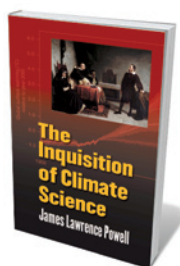
Genetics and cosmology owe their origins to two physicists, Max Delbrück and George Gamow, respectively. In this insightful double biography, theoretical physicist Gino Segrè portrays them as exemplary yet ordinary scientists who overcame personal struggles to work on big questions. He describes how Delbrück's flight from Nazi Germany turned him towards biology; and how Gamow's escape from Stalinist Russia to the United States led to him working on the hydrogen bomb and the formation of elements in the Big Bang.



### Future Science: Essays from the Cutting Edge

Edited by Max Brockman VINTAGE 272 pp. \$15.95 (2011)

In a sequel to *What's Next?* (Vintage, 2009), editor Max Brockman's latest collection of essays showcases the cutting-edge research of 19 leading young scientists — only a few of whom have written for a general audience before. Contributions include discussions of the biology of antiviral immunity by virologist William McEwan; the physical impact of social rejection by neuroscientist Naomi Eisenberger; the physics of infinity, discussed by physicist Anthony Aguirre; and the role of huge data sets in society by computer scientist Jon Kleinberg.



### The Inquisition of Climate Science

James Lawrence Powell COLUMBIA UNIVERSITY PRESS 272 pp. \$27.95 (2011)

Geologist James Powell exposes the tactics of climate-change deniers in his latest book. Arguing that the current attempt to undermine trust in climate research is the most egregious attack on science in history, he examines the movement and its protagonists. He points out the rhetorical tricks, lack of credentials and industry backing of many deniers, their unwillingness to present alternative theories, and historical precedents of resistance to scientific evidence.





JKM COLLECTION, NATL MUS, WILDLIFE ART/ESTATE OF CARL RUNGUIS

Caribou on the move as a storm rolls across the Rocky Mountains, by wildlife painter Carl Rungius (*Fall Storm*, 1935).

#### CONSERVATION

# How art is saving the West

**Anthony King** delights in a Wyoming exhibition focusing on the landscapes and wildlife in the migratory corridor between Yellowstone and Yukon.

Art has played a prominent part in conservation in North America. The paintings of Thomas Moran helped to inspire the US Congress to create Yellowstone, the world's first national park, in 1872. The exhibition *Yellowstone to Yukon*, now showing at the US National Museum of Wildlife Art in Jackson Hole, Wyoming, revisits this artistic tradition and engages visitors with the conservation issues surrounding one of the largest intact wildlife corridors on Earth.

The exhibition emphasizes how artists, well before conservationists, anticipated the idea of an ecosystem without borders running from Yellowstone to Yukon — a distance of some 3,200 kilometres. Harvey Locke, one of the founders of the Yellowstone to Yukon Conservation Initiative (Y2Y), is the driving force behind the show. He brought together art from the Wyoming museum and the Whyte Museum of the Canadian Rockies in Banff, Alberta, where the collection will move in 2012, to illustrate how artists at the end of the nineteenth century helped to establish the idea of wilderness and wildlife as valuable.

Sublime scenes on display include Albert Bierstadt's *Elk Grazing in the Wind River Country* (1861); John Fery's *Jackson Lake and the Tetons* (painted around 1900), depicting a scene that has since disappeared

**Yellowstone to Yukon: The Journey of Wildlife and Art**  
National Museum of Wildlife Art, Jackson Hole, Wyoming, USA.  
Until 14 August.

with the building of a dam; and *Great Falls of the Yellowstone*, painted around 1884 by Thomas Hill, one of the most prominent artists to follow Moran. All the big names are here. German-born Carl Rungius (1869–1959) is for many the Michelangelo of wildlife art. His impressionist canvasses depict wild bison (*The Last of the Buffalo*, 1900), moose and caribou (*Moose, Upper Ram River Valley* and *Fall Storm*, 1935), sheep (*In the Ogilvie Rockies*, around 1940) and landscapes (*Lake McArthur*, 1925).

Today, the region's riches are chronicled by contemporary painter Dwayne Harty, whom Locke asked to retrace Rungius's footsteps, and to document areas where the master had never set his easel. Harty travelled by four-wheel-drive vehicle across Yukon, rafted down the Nahanni River, flew to the isolated Ram Plateau by helicopter and rode on horseback for a month in the wilds of Muskwa-Kechika in British Columbia to paint animals large and small.

Locke is particularly concerned about the survival of large carnivores, especially grizzly

bears. Barriers are springing up across the wildlife corridor, notably along Highway 3, which cuts across the Rockies near the US–Canada border. Increased traffic and housing developments are steadily impeding migrations and isolating US wildlife populations from larger groups to the north. Y2Y is working to secure adjacent lands to maintain wildlife connectivity. Harty's artworks *Duck Pond Lake* and *Pileated Woodpecker* depict vital areas that remain intact.

When Harty revisited the scenes of some of Rungius' paintings, the vistas had changed little. Bighorn sheep are still grazing the meadows of Wilcox Pass, as they did 100 years ago. Harty painted mature rams close to a road in Jasper National Park, which was created in 1907 and is almost 2,000 square kilometres larger than Yellowstone.

A watercolour from the Whyte collection shows a deer stepping out in front of cars and cafes on an avenue in Banff, painted by Walter Joseph Phillips in 1947. Bronwyn Minton, a curator at the Wyoming museum, says that wildlife runs all over Jackson; moose show up wherever there is good willow. The view from the museum can include elk, bison, pronghorn and bald eagles. ■

**Anthony King** is a writer based in Dublin.  
e-mail: [anthonyjking@gmail.com](mailto:anthonyjking@gmail.com)

**NATURE.COM**  
For more on US environmentalism:  
[go.nature.com/npprrv](http://go.nature.com/npprrv)

# CORRESPONDENCE

## No tech gaps in *E. coli* outbreak

As president of the German Society for Hygiene and Microbiology, I object to Rainer Fislage's claim that the recent *Escherichia coli* outbreak exposed a technological gap in Germany's medical-microbiology infrastructure (*Nature* **475**, 174; 2011).

Many university and government institutes contributed to a huge, rapid nationwide effort to identify the infectious agent, using a combination of field epidemiology, molecular typing and high-throughput genomics. For example, the Institute for Hygiene at the University of Münster is part of a national network of microbiology reference centres and consulting laboratories appointed by the Robert Koch Institute in Berlin. Institute director Helge Karch and his team used molecular tools to identify the outbreak strain and determined its multilocus sequence type within 48 hours of receiving the first stool samples.

Five days later, they posted a polymerase chain reaction assay on their website ([www.ehec.org](http://www.ehec.org)) that enabled the scientific community and public-health laboratories to differentiate the outbreak clone from other enterohaemorrhagic *E. coli*. Four days after that, they submitted the first draft genome sequence of an outbreak isolate to the US National Center for Biotechnology Information (see [go.nature.com/tjdrth](http://go.nature.com/tjdrth) and A. Mellmann *et al.* *PLoS ONE* **6**, e22751; 2011). They have also published three in-depth analyses of the latest outbreak strain. Scientists at the Göttingen Genomics Laboratory also swiftly generated and analysed another draft sequence (E. Brzuszkiewicz *et al.* *Arch. Microbiol.* <http://dx.doi.org/10.1007/s00203-011-0725-6>; 2011).

There are lessons to be learned from the *E. coli* O104:H4 outbreak, and adjustments to be made to improve preparedness. But technology gaps were not a relevant problem in its handling.

**Sebastian Suerbaum** *German Society for Hygiene and Microbiology, and Hanover Medical School, Germany.*  
[suerbaum.sebastian@mh-hannover.de](mailto:suerbaum.sebastian@mh-hannover.de)

## Insurers could help address climate risks

Working with South Africa's largest short-term insurer, Santam, we investigated how communities should manage the increased risks associated with climate change. The global insurance industry has focused on refining the quantification, differentiation and pricing of the risk exposure of insured assets. Our findings call into question a sole reliance on this strategy (J. Nel *et al.* *CSIR/NRE/ECOS/2011/0063/B*; CSIR, 2011).

More than 80% of weather-related insurance claims in South Africa's southern Cape region were incurred in the past five years. Worse, climate models indicate that fire, flood and sea-storm risks in the area are set to increase significantly in the next 40 years. We found that a 1-in-75-year flood risk, as determined by rainfall data, could change to 1 in 45 years as a result of observed changes in land use. Risk is a dynamic and emergent property of nonlinear relationships between different risk drivers. In such fast-changing complex systems, static spatial differentiation of risk becomes less relevant.

The insurance industry should determine which assets share risk drivers and then encourage communities to manage these. Insurance can positively shape

societal behaviour — some previously reactive medical insurers have become proactive promoters of healthy living.

We also find that human-induced changes to the landscape can have an equal or greater effect on risk exposure than climate change. By actively managing and restoring the ecological buffering capacity of these landscapes, communities can significantly reduce current and future risks associated with climate change.

**Deon C. Nel** *WWF South Africa, Newlands, South Africa.*  
[dnel@wwf.org.za](mailto:dnel@wwf.org.za)

**Clifford Shearing** *University of Cape Town, Rondebosch, South Africa.*

**Belinda Reyers** *Council for Scientific and Industrial Research (CSIR), Stellenbosch, South Africa.*

## Admin burden is part of the job

Adam James voices a common frustration of senior scientists regarding their administrative burden (*Nature* **475**, 257; 2011). This viewpoint presumes a narrow definition of scientific practice and a rigid dichotomy between administrative and research tasks. Scientists should view administration as part of the research process, rather than eschewing it.

Scientific administration is a means of articulating science beyond the lab and is best performed by senior scientists with a wide-ranging knowledge of science and its contexts. It demands a sound knowledge of the science being practised, and improves as research networks expand. It helps to implement science at the level of the individual experiment as well as the discipline.

An enormous administrative effort went into institutionalizing molecular biology (N. C. Mullins *Minerva* **10**, 51–82; 1972). And as long ago as 1839, Charles Darwin

had to organize the logistics of his lengthy voyage on HMS *Beagle* and categorize his samples.

Science confers considerable pressures on its practitioners. Entrants to the profession would do well to recognize that managing these pressures entails balancing research and administration, not separating them.

**John N. Parker** *National Center for Ecological Analysis and Synthesis, Santa Barbara, California, and Arizona State University, Tempe, USA.*  
[parker@nceas.ucsb.edu](mailto:parker@nceas.ucsb.edu)

**Niki Vermeulen** *University of Vienna, Austria, and University of Manchester, UK.*

**Bart Penders** *Maastricht University, and Radboud University, the Netherlands.*

## China's grain relies on foreign fertilizer

China does need foreign help to feed itself (*Nature* **474**, 7; 2011) when it comes to other factors related to grain production, such as fertilizers.

The country consumes roughly one-third of the world's commercial fertilizers, and this is set to rise by 2–3% annually as its population increases to almost 1.5 billion.

China has been self-sufficient in nitrogen and phosphorus fertilizers since 1997 and 2006, respectively, but still imports about half of its potassium fertilizer. Its clean phosphate rock resource (almost 700 million tonnes) could run out in 30–40 years as demand continues to rise.

Without importing commercial fertilizers, energy and technology to produce grain more efficiently, China will be unable to meet its future grain needs.

**Chen Fang, Wan Kai-yuan, Liu Yi, Tao Yong** *Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China.*  
[fchen@ipni.ac.cn](mailto:fchen@ipni.ac.cn)



## Making mountains out of a moon

The Moon's cratered surface preserves the record of impacts that occurred during the late stages of its accretion. New simulations show that a collision with a companion moon may have formed the lunar farside highlands. [SEE LETTER P.69](#)

MARIA T. ZUBER

The Moon dramatically exemplifies the role that collisions played in shaping the solid planets, particularly during the earliest history of the Solar System. The Moon's very existence is believed to be a consequence of debris accreting in Earth's orbit after a Mars-sized body collided with Earth shortly after it formed<sup>1,2</sup> — a theory known as the giant-impact hypothesis. Lava-filled basins on the lunar nearside, and the massive South Pole–Aitken basin on the farside, mark the locations of major asteroid-sized impacts after the Moon accreted, melted and cooled. And thousands of craters all over the surface represent 'scars' of abundant smaller impacts, mostly accumulated during the waning stage of planet formation. The common attribute of the geological record of impact events is that they produced 'holes' — craters and basins. Now Jutzi and Asphaug (page 69 of this issue<sup>3</sup>) present intriguing simulations suggesting

that a collision by a lunar companion that was created in the Moon-forming impact produced the extensive mountainous terrain on the lunar farside, rather than a crater or basin (Fig. 1).

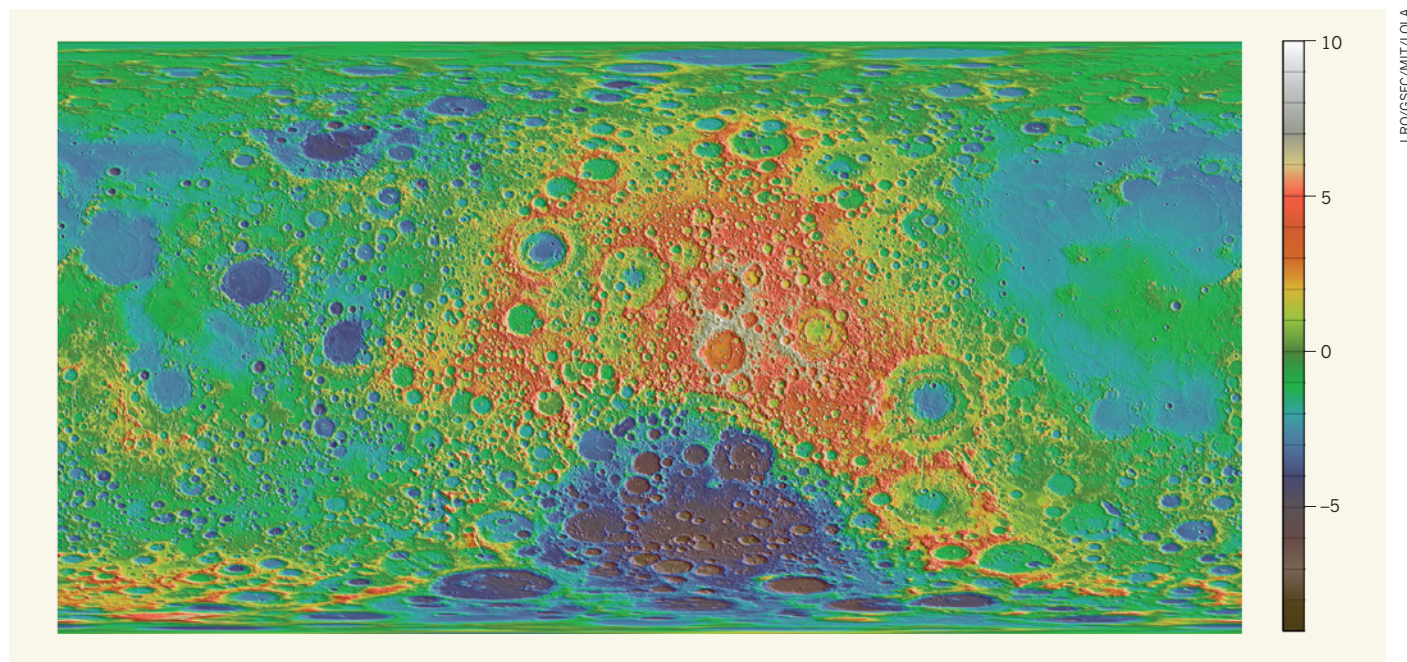
The idea that multiple moons could have formed after the giant impact had been demonstrated<sup>4</sup> by previous two-dimensional simulations<sup>4</sup> of the dynamics of the Earth-encircling debris disk that was generated by the impact. Calculations<sup>4</sup> showed that perturbations caused by the largest moonlet (our Moon) as it receded from Earth effectively removed material from the disk inside the Moon's orbit; a companion in this region would probably not have survived long.

The situation could have been different, however, for a companion at Earth–Moon Trojan points — locations ahead of and behind the orbit of the Moon that mark positions where the gravitational attraction of the Earth and Moon balance. Objects at these points are more stable than those at other parts of the Earth–Moon system, and could survive for up

to tens of millions of years. At such time after accretion, the lunar crust would have been in place, probably consisting of a global 'magma ocean' with a thin, overlying crystallized crust<sup>5</sup>.

So a lunar companion could conceivably survive long enough for its eventual collision to occur after the Moon was fully accreted and partially cooled. But how could the impact make mountains rather than a hole? To do so would require special conditions, which Jutzi and Asphaug elucidated using calculations performed by a computer code<sup>6</sup> of the kind used to study the giant impact. The authors' simulation considered the effect of planet-scale collisions by taking into account impact forces and gravity, as well as the deformational properties of relevant geological materials. The authors found that, to create elevated terrain of the volume and position of the farside highlands, a companion moon with a diameter about a third that of the current Moon would have been needed.

The companion, because it would have



**Figure 1 | The Moon's enigmatic farside highlands.** The image shown is a cylindrical projection of lunar topography obtained from the Lunar Orbiter Laser Altimeter<sup>11</sup> centred on the farside highlands, shown in red/yellow. Elevations are given in kilometres relative to a sphere of 1,737.4 km. The South Pole–Aitken basin is shown in dark colours below the highlands, and the lunar nearside is at the left and right. Jutzi and Asphaug's simulations<sup>3</sup> suggest that a collision by a lunar companion may have formed the farside highlands.

formed from the same debris disk as the Moon, would have had a similar composition, but would have solidified before the Moon did, so the age at which its rocks crystallized would be older than for lunar rocks. The authors deduced that the companion would need to have impacted the Moon at a velocity lower than the speed of sound, and that the volume of the lunar crust excavated must have been much less than the volume of the impactor. In addition, most of the material from the impacting companion must have remained near the impact point. This is in contrast to impacts that produce craters and basins, where the impactor exceeds the sound speed. In these hypervelocity events, material is excavated and ejected away from the impact point, and target rocks are melted and vaporized in amounts that scale with the kinetic energy of the event.

In the highlands-forming collision, the addition of material from the companion would have thickened the lunar crust, and such thickening is observed from analyses of lunar topography and gravity<sup>7</sup>. The collision would also have preferentially redistributed the underlying magma ocean to the nearside of the Moon, consistent with the distribution of heat-producing elements observed<sup>8</sup> by the Lunar Prospector spacecraft.

The origin of the farside highlands has been a topic of speculation since the first global measurement<sup>9</sup> of the Moon's shape. Ideas have included piling up of material excavated from the adjacent South Pole–Aitken impact<sup>9</sup> basin, variations in tidal heating<sup>10</sup>, and a variety of other processes that produce spatial heterogeneities, for example shallow melting, crustal crystallization, interior dynamics and impact bombardment. The challenge is that the distributions of the Moon's elevation and crustal thickness calculated from orbital observations broadly fit sinusoids, and because numerous processes in nature can produce such a pattern, the ability to fit does not in itself distinguish between different mechanisms. Consequently, the current study<sup>3</sup> demonstrates plausibility rather than proof. Because the accreted companion would have formed earlier than the Moon, demonstrating that the farside highlands have older crystallization ages would bolster support for this idea. However, that would require returning samples of the farside highlands to Earth to calculate the absolute ages.

More feasible in the near term will be to match simulations such as those carried out by the authors to the detailed internal structure of the Moon that will be obtained by topography (Fig. 1) from the Lunar Reconnaissance Orbiter mission and by high-resolution gravity mapping, which will be obtained next year by the Gravity Recovery and Interior Laboratory (GRAIL) mission. In any case, Jutzi and Asphaug raise<sup>3</sup> the legitimate possibility that, after the giant impact, our Earth perhaps fleetingly possessed more than one moon. Furthermore, significant remnants of this

long-departed member of the Earth–Moon collisional family may be preserved today on the lunar farside. ■

**Maria T. Zuber** is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: zuber@mit.edu

1. Cameron, A. G. W. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 609–620 (Lunar Planet. Inst., 1986).
2. Canup, R. M. & Esposito, L. W. *Icarus*

- 119, 427–446 (1996).
3. Jutzi, M. & Asphaug, E. *Nature* **476**, 69–72 (2011).
4. Canup, R. M., Levison, H. F. & Stewart, G. R. *Astron. J.* **117**, 603–620 (1999).
5. Warren, P. H. *Annu. Rev. Earth Planet. Sci.* **13**, 201–240 (1985).
6. Benz, W. & Asphaug, E. *Comput. Phys. Commun.* **87**, 253–265 (1995).
7. Neumann, G. A., Zuber, M. T., Smith, D. E. & Lemoine, F. G. *J. Geophys. Res.* doi:10.1029/96JE01246 (1996).
8. Lawrence, D. J. *et al. Science* **281**, 1484–1489 (1998).
9. Zuber, M. T., Smith, D. E., Lemoine, F. G. & Neumann, G. A. *Science* **266**, 1839–1843 (1994).
10. Garrick-Bethell, I., Nimmo, F. & Wieczorek, M. A. *Science* **330**, 949–951 (2010).
11. Smith, D. E. *et al. Geophys. Res. Lett.* doi:10.1029/2010GL043751 (2010).

## HIGH-TEMPERATURE SUPERCONDUCTIVITY

# The great quantum conundrum

**Twenty-five years on from its discovery, high-temperature superconductivity remains without a satisfactory explanation. The latest studies on the electronic phase diagram of copper oxide compounds reveal why this is so. SEE LETTER P.73**

PAUL MICHAEL GRANT

Some years ago, I lectured at a National Science Foundation summer workshop for high-school physics teachers. My subject was superconductivity. One of my co-instructors was Robert Laughlin. Scrawled across the top of Bob's first projector slide was the phrase, 'The Theory of Everything', and I thought, 'Oh, boy, here we go, the standard model of particle physics — again'. But underneath the title, he had written instead the many-body Schrödinger equation, summed over all the interactions between electrons and nuclei, and thus containing, once electron spin is included, the complete chemistry and physics of ordinary, terrestrial matter.

Of course, the devil is always in the details, in this case the enormous summation over particle coordinates that is required to achieve a scale of, say, Avogadro's number. From this summation emerge life, the climate, smartphones ... and high-temperature (high- $T_C$ ) superconductivity. And it is on this last that Jin *et al.*<sup>1</sup> (page 73 of this issue) and He *et al.*<sup>2</sup> (in an earlier study published in *Science*) make the latest effort to illuminate qualitatively the microscopic origins. They do this by attempting to unravel the enigmas of the electronic phase diagram of materials known as copper oxide perovskites. Within this phase diagram (Fig. 1) reside several quantum states, characterized by one or more 'quantum critical points', in rough analogy to the classical critical points characterizing the separation of the gas, liquid and solid states of macroscopic matter. How the various phases in the

electronic phase diagram compete or cooperate in generating the emergent superconducting state constitutes what I term the great quantum conundrum.

The physics of the '3d-transition-metal' monoxides, such as copper monoxide, is fascinating. By conventional energy-band theory, these compounds should all be metals. But they are not: they are a form of non-conductor called an antiferromagnetic Mott insulator, in which neighbouring cations (metal atoms) contain one or more opposite spins. This was the great quantum conundrum of my generation, which was formally cleared up by John Hubbard<sup>3</sup>.

Things get even more interesting when we consider what happens if charge from some external 'dopant' source is added to (with electrons) or subtracted from (with holes) the material's copper–oxygen energy bands. Researchers had tried this approach with other transition-metal oxide compounds, but found nothing much. Then Georg Bednorz and Alex Müller ran across barium-doped, layered lanthanum copper oxide and discovered high- $T_C$  superconductivity. The presence of barium introduced holes into the copper oxide bands, destroying long-range antiferromagnetic order and creating — depending on the temperature and charge-carrier concentration — either normal conductivity or superconductivity. Similar behaviour was also later found when electron-donating cations were employed. As Phil Anderson cogently pointed out in 1986, once the Bednorz–Müller findings had received wide notice, it was "unlikely an accident that the original host material for high



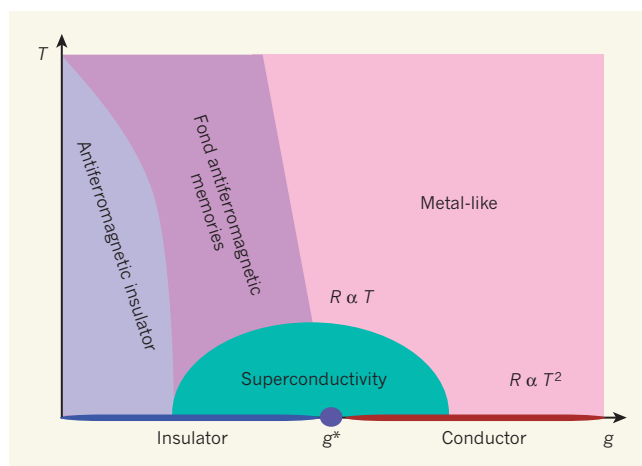
temperature superconductivity was an antiferromagnetic insulator”.

The explosion of high- $T_C$  superconductors wrought by modern materials science has meanwhile produced a plethora of phase diagrams of temperature versus local charge density (induced by doping and/or pressure), typified by Figure 1. One common feature of these diagrams is a  $T = 0$  line, plotted as a function of local charge density, which begins at low values describing a doped antiferromagnetic insulator and eventually reaches a quantum critical point, beyond which conductivity emerges.

The latest efforts at improving these qualitative descriptions are reported by Jin *et al.*<sup>1</sup> and by He *et al.*<sup>2</sup> Jin and colleagues focused on the electron-doped copper oxides, whereas He *et al.* addressed their hole-doped counterparts. However, the two groups approach their subject from different directions and arrive at somewhat different points of view.

It has long been speculated that some remnant of magnetic order remains behind when the nascent copper oxides are doped into a conducting or a superconducting state — I like to term this the fond memories of antiferromagnetism. In principle, such memories can be detected experimentally and may be linked to a phenomenon known as the pseudogap. The pseudogap, whose existence is revealed by anomalies in photoemission experiments, has been observed by a number of groups, especially in the hole-doped copper oxides. It is characterized by a transition temperature beyond which a metallic-like state emerges, and lies on top of the dome-shaped superconducting region of the phase diagram (Fig. 1). A central issue, however, is whether the pseudogap has a competitive or collaborative role in engendering high- $T_C$  superconductivity. He *et al.*<sup>2</sup> claim the latter, pointing out that the energy of the pseudogap approaches that of the superconducting gap, as both converge in temperature towards the summit of the superconducting dome.

For electron-doped copper oxides, the case for the demarcation of a pseudogap phase is less clear than for hole-doped materials. Therefore, Jin *et al.*<sup>1</sup> focused on a common feature seen in both systems — that the resistance of the normal conducting state shows a subquadratic or quasi-linear dependence on temperature that is observed in broad regions of local charge density above the superconducting dome. I recall hearing Phil Anderson remark that the linear dependence of the normal conducting-state resistance on temperature was as unusual as the appearance of



**Figure 1 | Generic electronic phase diagram of the copper oxide superconductors.** The diagram shows the major observed phases as a function of temperature ( $T$ ) and local charge density ( $g$ ), which is introduced by either doping or pressure: from left to right, an antiferromagnetic insulator, a state in which carriers start to move but ‘fondly remember’ their antiferromagnetic origins (this is where the pseudogap phenomenon is located), and a metal-like phase whose resistance ( $R$ ) varies linearly to quadratically with temperature as  $g$  increases. These phases sit above a ‘superconducting dome’ whose boundary remains a subject of intense debate and which is vital to understanding the origin of high-temperature superconductivity. At zero temperature, as  $g$  increases and reaches a ‘quantum critical point’ ( $g^*$ ), the copper oxides cease to be insulators and become conductors. The studies by Jin *et al.*<sup>1</sup> and by He *et al.*<sup>2</sup> attempt to provide insights towards a better understanding of this phase diagram.

superconductivity. This temperature dependence in ordinary metals at low temperature goes as  $T^2$ .

Jin *et al.* attribute this linear temperature dependence to ‘spin-fluctuation’ scattering of carriers, and speculate that this scattering is also responsible for the carrier-pairing mechanism underlying superconductivity in the layered copper oxides. They point to an analogous behaviour in the layered organic Bechgaard salts, compounds in which the local charge density can be varied by applying hydrostatic pressure, rather than by direct doping. This results in a superconducting state at around 10 kilobar up to a  $T_C$  of about 1.2 kelvin, and a subquadratic temperature dependence in the normal conducting-state resistance at higher temperatures.

So, there is certainly circumstantial evidence that high- $T_C$  is mediated by spin fluctuations, but is it compelling? Or, put another way, there does seem to be a ‘smoking gun’ here, but what was it that pulled the trigger? What exactly is a spin fluctuation? One might think that there should even be a model at the microscopic scale from which one could calculate the coefficient of proportionality in the linear-temperature formula outlined by Jin *et al.*<sup>1</sup>. However, a search of the reference list in the papers by He *et al.* and Jin *et al.* revealed no source for such a model, nor did perusal of several recent texts on the theory of unconventional superconductors. What the theoretical framework of superconductivity in

the layered copper oxides presently lacks is an equivalent to the Eliashberg–McMillan tool<sup>4,5</sup>, which can be used to ‘post-dict’ the transition temperature of simple low- $T_C$  superconductors using an electron–lattice vibration (phonon) interaction. This can be calculated directly for, say, aluminium or niobium using a computational method known as density functional theory, but same methodology could, in principle, be used to treat spin-fluctuation scattering.

Having mentioned electron–phonon coupling, I should point out that most, if not all, of the layered copper oxides display an isotope effect. This occurs when one of the elements is exchanged for a lighter or a heavier one, which has a different number of neutrons in its nucleus. The substitution changes the ‘strength’ of the electron–phonon interaction, and causes a shift in  $T_C$ . In addition, in those instances when it has been measured, the neutron diffraction pattern shows anomalies near  $T_C$ . Both of these effects strongly suggest that phonons are somehow involved in whatever carrier-pairing physics applies to high- $T_C$  superconductors.

Let us return to the 3d-transition-metal monoxides. Most of the heavier ones (manganese, iron, cobalt, nickel, copper, zinc) can be found in ‘everything Laughlin’ in cubic rock-salt form. The one exception is copper monoxide, which emerges as the mineral tenorite, and which has a highly distorted crystal structure. I have applied<sup>6</sup> density functional theory to investigate the stability of ‘cubic rock-salt’ copper monoxide, and found it to be extremely unstable, easily undergoing uniaxial deformation. Such instability results from the highly degenerate copper–oxygen bond, in effect reflecting an unusually strong electron–phonon interaction, much greater than is found in the other transition-metal monoxides. It is just this kind of interaction, originally postulated by Alex Müller, that he concluded might lead to unusually high-temperature superconductivity in the copper oxides. Was he right, after all?

So, does high- $T_C$  superconductivity emerge from ‘everything Laughlin’ because of ‘lattice shakes’ or ‘bouncing spins’? And do these effects compete or cooperate? I suspect the latter. At the end of the day, one should try to look beyond the current fashions of jargon and arguments-by-analogy, and move towards a formalism for high- $T_C$  carrier pairing that can be used to compute  $T_C$  on a more-or-less first-principles basis, much as now can be done for low- $T_C$  superconductors. A newly published paper by Le Tacon and colleagues<sup>7</sup>, which

attempts a model calculation of spin-fluctuation-mediated pairing, provides an encouraging start. In this context, it may be pertinent to quote the response to David Mermin's observation<sup>8</sup>, when questioning the physics faculty while a graduate student at Harvard, about the weirdness of that first quantum conundrum — entanglement. He was advised to simply “shut up and calculate”. ■

**Paul Michael Grant** is at W2AGZ Technologies and is an emeritus research staff member at the IBM Almaden Research Center, San Jose, California 95120, USA. e-mail: pmpgrant@w2agz.com

1. Jin, K., Butch, N. P., Kirshenbaum, K., Paglione, J. & Greene, R. L. *Nature* **476**, 73–79 (2011).
2. He, R.-U. *et al. Science* **331**, 1579–1583 (2011).

3. Hubbard, J. *Proc. R. Soc. Lond. A* **276**, 238–257 (1963).
4. Eliashberg, G. M. *Sov. Phys. JETP* **11**, 696–702 (1960).
5. McMillan, W. L. *Phys. Rev.* **167**, 331–344 (1968).
6. Grant, P. M. *J. Phys. Conf. Ser.* **129**, 012042 (2008).
7. Le Tacon, M. *et al. Nature Phys.* <http://dx.doi.org/10.1038/nphys2041> (2011).
8. Mermin, N. D. *Phys. Today* **57**(5), 10–11 (2004).

## ANTHROPOLOGY

# Shades of the savannah

**Whether African savannahs had an impact on the evolution of our early ancestors has been a matter of debate. A study of carbon isotopes from ancient soils provides fresh clues. SEE ARTICLE P.51**

CRAIG S. FEIBEL

We are creatures of the African savannah. Through some six million years of evolution, members of our branch of upright, bipedal primates — the hominins — have adapted to the prevailing conditions of savannah life. Writing on page 51 of this issue, Cerling *et al.*<sup>1</sup> add a new dimension to our understanding of what early savannahs looked like, and how they changed through time, ultimately shaping our evolutionary course.

Although there is a rich trove of fossils recording the evolution of the hominins, as well as abundant stone artefacts that document their cultural awakenings, the environment that shaped them over time has not been preserved. Sediments reflect important details of depositional settings, and proxy evidence from lakes and oceans has contributed

to a growing understanding of past climates. But the component of the savannah setting that is most difficult to reconstruct forms the very core of the savannah concept — the vegetation that clothed the ancient landscape. In this context, the rarity of plant fossils and debates surrounding the terminology used to describe the diverse manifestations of the mix of habitats within a local landscape, known as the savannah mosaic, have clouded our understanding of the conditions and changes that shaped our evolutionary path<sup>2,3</sup>.

An underlying tenet to most discussions of human origins is the savannah hypothesis. In its broadest manifestation, it is the idea that environmental change transformed forests of the Miocene Epoch — some 23 million to 5 million years ago — into the savannahs of Plio-Pleistocene times (the past 5 million years), and that this shift led to upright bipedalism, increased brain mass and the

suite of other characteristics that make us human.

However, in recent years, further site discoveries have challenged this classical view. For instance, the idea that Miocene Africa was home to unbroken forest has begun to fragment, with evidence<sup>4</sup> of true desert in the Sahara. Meanwhile, suggestions of more wooded conditions during the Plio-Pleistocene clouded the other side of the transformation<sup>5</sup>. At the same time, semantic arguments about the meaning of the term savannah — encompassing almost everything between desert and forest in Africa — have crippled much of the discussion.

Cerling and colleagues' work<sup>1</sup> enlists isotopic data to tackle this problem. The authors deftly sidestep ambiguities about the exact meaning and implications of the term savannah by focusing on a discrete, quantifiable characteristic of the savannah mosaic — the percentage of woody cover. Using a data set of some 1,300 samples from ancient soils, the team tracked the woody component, and hence the percentage of trees and available shade, through some six million years of savannah history. They drew the samples from two of the richest fossil regions of East Africa — the Awash and Omo-Turkana basins (Fig. 1).

The results show less than 40% of woody cover for most of the study period; thus relatively open habitats were common from the

T. CERLING



K. GARRETT/NATIONAL GEOGRAPHIC STOCK

**Figure 1 | Sampling sites.** In their efforts to identify the importance of a savannah habitat in human evolution, Cerling *et al.*<sup>1</sup> collected soil samples from the Awash (left) and Omo-Turkana (right) basins in eastern Africa.



early ancestry of the hominins. The details of variation through time are equally striking. Whereas more open habitats prevailed for the creatures that some regard as the earliest bipeds, woodier habitats expanded throughout the early range of *Australopithecus* — a more broadly accepted early biped.

Intriguing questions remain. The material Cerling *et al.* used for most of their analyses is soil carbonate — minerals precipitated within the soil and which integrated the isotopic signal of carbon dioxide found there. These nodules form primarily where evaporation greatly exceeds precipitation. But many ancient

soils lack these carbonate nodules, reflecting wetter times or moister areas in the landscape. How well this proxy reflects the overall temporal and spatial variability in ancient vegetation is unclear.

Placing our ancient ancestors within particular habitats on these landscapes is another wild card. Most of the fossils we have were found where an organism died, or where its bones were transported to, and not necessarily where it lived. Studies focusing on spatial variability across landscapes, and intensive collecting protocols linking fossil assemblages and communities to specific habitats, are under

way to address these issues. Nonetheless, Cerling and co-workers' data<sup>1</sup> go a long way to reconstructing a fundamental control on the characteristics and variation of the ancient habitats available to our ancestors. ■

**Craig S. Feibel** is in the Department of Earth and Planetary Sciences, Rutgers University, Piscataway, New Jersey 08854-8066, USA.  
e-mail: feibel@rci.rutgers.edu

1. Cerling, T. E. *et al.* *Nature* **476**, 51–56 (2011).
2. White, T. D. *et al.* *Science* **326**, 75–86 (2009).
3. Cerling, T. E. *et al.* *Science* **328**, 1105 (2010).
4. Schuster, M. *et al.* *Science* **311**, 821 (2006).

respond to similar temperatures. Trigeminal nerves also receive sensory input from the head and face in bats and snakes. But in bats, they innervate detectors not in pits<sup>1</sup> but on the upper lip and modified noseleaf<sup>3</sup> (Fig. 1).

How do these differences affect the animals' hunting behaviour? Pit vipers often prey on small mammals and can better detect hot spots at night. A good example of the importance of thermal cues to these snakes is the defensive behaviour of California ground squirrels<sup>4</sup>. When confronted by a pit viper such as a Pacific rattlesnake, ground squirrels flag their tails to distract the predator. This display has a thermal component that is missing when the predator is a gopher snake — a species that lacks thermosensation. Like pit vipers, some species of python and boa also use infrared sensors in facial pits to detect warm-blooded prey and to guide their strikes even in the absence of visual cues<sup>5</sup>.

Vampire bats can detect a heat source from about 20 cm (ref. 3), and probably use this proximal cue to find hot spots on their prey — often areas that are not covered with fur or feathers. Repeated attacks on the same cattle suggest that, together, the vampire's memory and the prey's breathing sounds<sup>6</sup> provide distal cues that allow the bats to locate a sleeping target<sup>7</sup>. As yet, there are no data to suggest defences against thermoperception by vampire bats similar to those of California ground squirrels against vipers. However, it is important to note that the bats take only about two tablespoonfuls of blood (25 ml), and so, unlike those of pit vipers, vampire attacks are not fatal. Nonetheless, there is evidence that some prey develop antibodies to the anticoagulants in the vampire's saliva<sup>8</sup>, directly affecting the bats' feeding time.

In vampire bats, the organization of the *Trpv1* gene seems to

## NEUROSCIENCE

# Heat-thirsty bats

**Vampire bats sense infrared radiation to locate places where blood flows close to their prey's skin. At a molecular level, this ability is underpinned by the intricate redesign of an ion channel on facial nerves. [SEE LETTER P.88](#)**

M. BROCK FENTON

The ability to sense heat or cold is vital: it allows animals to detect, and so avoid, debilitating or fatal temperatures. But there is more to thermosensation than keeping one's cool (or heat). On page 88 of this issue, Gracheva and colleagues<sup>1</sup> report on modifications to the facial nerves that allow the common vampire bat (*Desmodus rotundus*) to detect infrared thermal radiation associated with hot spots — areas where blood flows close to the skin of the bats' prey — so that they can efficiently access the blood they eat.

This paper complements earlier work<sup>2</sup> on infrared sensors in snakes, including some boas, pythons and pit vipers, which also use radiation in this range to locate food. The endings of somatosensory trigeminal nerves in the pits on the pit viper's face detect infrared radiation. Specifically, pit vipers detect infrared radiation through TRPA1 — a cell-surface ion channel that is usually heat insensitive — on these nerves. The heat sensors of pit vipers are more sensitive than those of pythons or boas, indicating independent evolution of thermosensation.

Gracheva *et al.*<sup>1</sup> show that vampire bats sense infrared radiation through a different ion channel, TRPV1, which is used by other mammals to detect noxious heat (> 43 °C). In vampire bats, however, the

TRPV1 activation threshold is lowered to about 30 °C through alternative splicing of its gene transcripts, which results in truncation of the channel's carboxy-terminal domain. The truncated version of the protein is not expressed in the bats' dorsal root ganglia, which innervate the spinal cord. Instead, it occurs only in their facial (trigeminal) nerves, where ion channels in both vipers and bats usually



**Figure 1 | Vampire bat.** Infrared sensors are located on the bat's upper lip and modified noseleaf.

M. B. FENTON

be characteristic of the Laurasiatheria, the superorder of mammals that includes bats (order Chiroptera), as well as several other orders. Chiroptera is further divided into two suborders, and Gracheva *et al.*<sup>1</sup> show that *Trpv1* occurs in both. The authors call these suborders microbats (Microchiroptera) and megabats (Megachiroptera). However, the two bat suborders are now called Yinpterochiroptera and Yangochiroptera<sup>9</sup>, and the split of species between the two is not the same as for microbats and megabats. Gracheva and co-workers' analysis of the *Trpv1* gene in bats does not alter the current view<sup>9</sup> of bat phylogeny.

Gracheva *et al.*<sup>1,3</sup> have placed infrared detection high on the list of astonishing discoveries about the perceptual abilities of animals: it seems to have evolved in parallel within two snake lineages, and converged with the appearance in vampire bats. Their data, together with other recent findings, also enrich our knowledge of the sensory world of bats. Previous work suggested, for instance, that bats' wing membranes are as sensitive to touch as our fingertips<sup>10</sup>. And tactile receptors associated with sensory hairs on the bat wings are

known to monitor flight speed and air flow<sup>11</sup>. The perceptual world of bats undoubtedly has many more intriguing secrets yet to be discovered. ■

**M. Brock Fenton** is in the Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada. e-mail: bfenton@uwo.ca

1. Gracheva, E. O. *et al.* *Nature* **476**, 88–91 (2011).
2. Gracheva, E. O. *et al.* *Nature* **464**, 1006–1011 (2010).
3. K rten, L., Schmidt, U. & Sch fer, K. *Naturwissenschaften* **71**, 327–328 (1984).
4. Rundus, A. S., Owings, D. H., Joshi, S. S., Chinn, E. & Giannini, N. *Proc. Natl Acad. Sci. USA* **104**, 14372–14376 (2007).
5. Grace, M. S. & Woodward, O. M. *Brain Res.* **919**, 250–258 (2001).
6. Gr ger, U. & Wiegube, L. *BMC Biol.* **4**, 18 (2006).
7. Schmidt, U., Schlegel, P., Schweizer, H. & Neuweiler, G. *J. Comp. Physiol. A* **168**, 45–51 (1991).
8. Delpietro, H. A. & Russo, R. G. *J. Mammal.* **90**, 1132–1138 (2009).
9. Teeling, E. C. *Trends Ecol. Evol.* **24**, 351–354 (2009).
10. Chadha, M., Moss, C. F. & Sterbing-D'Angelo, S. J. *J. Comp. Physiol. A* **197**, 89–96 (2011).
11. Sterbing-D'Angelo, S. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 11291–11296 (2011).

## OCEANOGRAPHY

## Forecasting the rain ratio

**Marine algae known as coccolithophores produce much of the ocean's calcium carbonate. A large survey reveals how these organisms' calcification processes and species distribution change in response to carbon dioxide levels. SEE LETTER P.80**

DAVID A. HUTCHINS

**C**occolithophores are humble marine phytoplankton that are the subject of a simmering controversy in marine science concerning their response to ocean acidification. On page 80 of this issue, Beaufort *et al.*<sup>1</sup> report a finding that should help settle the matter: coccolithophores produce thinner calcium carbonate shells as oceans become more acidic. But the mechanisms involved, and an unexpected exception to the general rule, may surprise those studying global change.

Over the past 220 million years or so, coccolithophores have performed a unique dual function in the ocean's carbon cycle. Like all phytoplankton, coccolithophores make their living by converting dissolved inorganic carbon in sea water into organic carbon through photosynthesis. But they also have a singular ability to use dissolved inorganic carbon to produce a mineral shell consisting

of coccoliths, overlapping plates of calcium carbonate ( $\text{CaCO}_3$ ). Although the alga's long evolutionary history spans several major fluctuations in atmospheric carbon dioxide content, predicting the responses of coccolithophore calcification to ocean acidification — the anthropogenic enrichment of the modern ocean with  $\text{CO}_2$  — has been anything but straightforward.

The question of how coccolithophore calcification will respond to future high  $\text{CO}_2$  conditions has big implications for the ocean's carbon cycle, and perhaps also for global climate. The ratio of  $\text{CaCO}_3$  to organic carbon in the continuous 'rain' of biogenic particles that sink down from the ocean's surface (the 'rain ratio'<sup>2</sup>) is a key factor in carbon biogeochemical models, for several reasons<sup>3</sup>. One of the most important is that, in contrast to the photosynthetic production of organic carbon, which consumes  $\text{CO}_2$ , the calcification reaction produces  $\text{CO}_2$  by converting two bicarbonate ions ( $\text{HCO}_3^-$ ) into one  $\text{CaCO}_3$



## 50 Years Ago

It should not be deduced from this that, his scientific training and proclivities apart, the good scientist of to-day is ignorant about or, even worse, unaware of, other branches of man's culture. In fact, the contrary is true; indeed, so far as the community of scientists is concerned, the so-called 'two cultures' (fashionably ascribed to C. P. Snow) scarcely exist. Many scientists are well read outside their own discipline, sometimes still within the ambit of science, but more often well beyond it — in philosophy, history, art, music, the theatre, literature, in fact, in the humanities generally. (Good scientists seldom make good politicians, which is probably understandable.) ... In short, it is high time that the general opinion, still very extant, that the man of science is so wrapped up in his scientific literature and so confined to his laboratory that, apart from his calling, he is culturally unbalanced, be challenged. Indeed, the shoe is on the other foot; it is the student of, and savant among, the humanities, art, music and non-scientific literature (especially fiction) who are — and are often proud to admit they are — quite ignorant of science and its now 'jet-propelled' progress.

From *Nature* 5 August 1961

## 100 Years Ago

'The Birmingham meeting of the British Medical Association' — According to Dr. Provan Cathcart, the quality, and not the quantity, of the protein is the important matter physiologically, for the nearer the composition as regards the constituent amino-acids approaches that of the tissue-protein of the animal being fed, the less will there be of nitrogenous waste from that animal. Thus dogs wasted less nitrogen when fed on dog-flesh than on any other kind of protein.

From *Nature* 3 August 1911



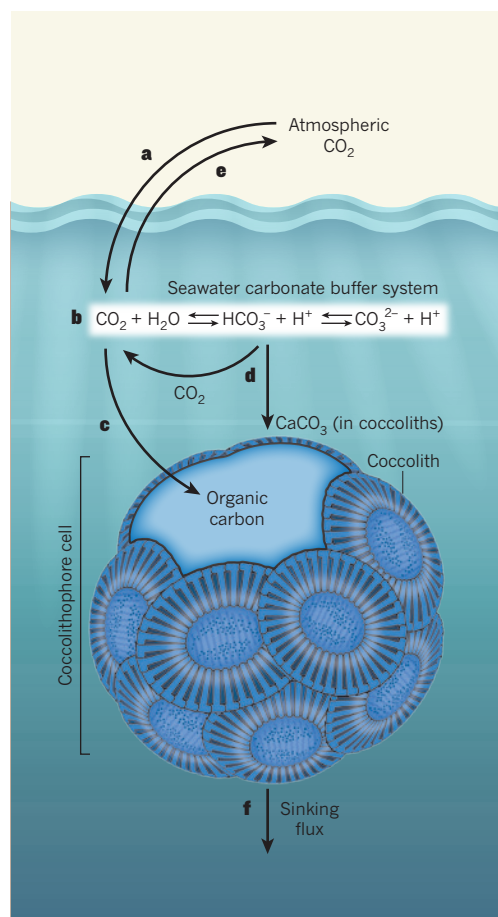
and one CO<sub>2</sub> molecule (Fig. 1). Thus, the concentration of CO<sub>2</sub> in sea water is sensitive to changes in the rain ratio, which is a proxy for the amount of calcification versus photosynthesis occurring in the ocean.

The link between calcification and fluxes of climate-altering CO<sub>2</sub> prompted experiments that found that the production of CaCO<sub>3</sub> was reduced in coccolithophores growing at elevated CO<sub>2</sub> levels<sup>4–6</sup>. Given that future ocean CO<sub>2</sub> concentrations are expected to be high, concomitant reductions in calcification would lower the rain ratio, potentially helping to counter the rise in atmospheric CO<sub>2</sub> concentrations. But just as marine scientists were becoming comfortable with this emerging ocean-acidification model, other papers<sup>7,8</sup> stirred the pot by reporting that high CO<sub>2</sub> levels increase the amount of CaCO<sub>3</sub> produced by coccolithophore cells.

Beaufort *et al.*<sup>1</sup> have now boldly charged into the resulting confusion and dissension. Unlike the previous studies, their work did not manipulate CO<sub>2</sub> levels in cultures<sup>4,6</sup> or in natural communities<sup>9,10</sup> of coccolithophores. Instead, they used image-analysis techniques to determine the masses of more than half a million individual coccoliths from hundreds of modern surface seawater samples and from ancient marine sediment cores, collected from all over the world. They also measured the corresponding concentrations of dissolved inorganic carbon in the modern seawater samples, or calculated these concentrations for the sediment cores using accepted palaeoceanographic proxies.

Their findings are unequivocal: as CO<sub>2</sub> concentration increases, coccolith mass declines in a more or less linear fashion. This relationship holds up regardless of the large local variations in seawater CO<sub>2</sub> concentrations found in today's oceans, and it also holds up over long-term temporal fluctuations in atmospheric CO<sub>2</sub>, such as those that have occurred over glacial–interglacial cycles. The results seem to offer solid support for the hypothesis that coccolithophore cells will be less calcified in the future acidified ocean. But there is another twist to the story.

Beaufort *et al.* point out<sup>1</sup> that the variations in coccolith mass measured in their study are much larger than the decreases in cellular CaCO<sub>3</sub> typically observed when single species of coccolithophores are grown in culture at high CO<sub>2</sub> concentrations. In fact, much of the coccolith-mass variability they recorded was apparently the result of taxonomic shifts in the coccolithophore community, rather than the result of reduced calcification within individual species. As levels of CO<sub>2</sub> in sea water increase, assemblages of the algae progressively shift away from larger, heavily calcified species and towards smaller, lightly calcified ones.



**Figure 1 | Coccolithophore carbon chemistry.**

**a, b,** When atmospheric CO<sub>2</sub> enters the sea surface (**a**), it undergoes a series of reversible chemical reactions known as the seawater carbonate buffer system (**b**), which releases protons (H<sup>+</sup>) that acidify the sea water. **c,** Coccolithophores and other algae assimilate CO<sub>2</sub> to produce organic carbon through photosynthesis. **d,** Coccolithophores also perform calcification reactions, in which two bicarbonate ions (HCO<sub>3</sub><sup>-</sup>) are converted into one calcium carbonate (CaCO<sub>3</sub>) and one CO<sub>2</sub> molecule. The CaCO<sub>3</sub> is incorporated into coccoliths in the algal shell. The CO<sub>2</sub> from calcification is released, and can either contribute to ocean acidification or degas back to the atmosphere (**e**), contributing to global warming. **f,** Biogenic particles from coccolithophores and other phytoplankton sink from the ocean surface. The ratio of CaCO<sub>3</sub> to organic carbon in this 'rain' of biogenic particles is a critical parameter in the marine carbon cycle. Beaufort *et al.*<sup>1</sup> show that coccolithophores produce less calcium carbonate at higher seawater concentrations of CO<sub>2</sub>.

This trend occurs even within species, so that robustly calcified strains or morphotypes are replaced by more delicately calcified ones as CO<sub>2</sub> levels rise. The authors' results therefore seem to imply that seawater carbonate chemistry is a strong selective force determining the taxonomic composition of coccolithophore communities.

So does this mean that the previously reported observations of increased cellular calcification in cultured coccolithophores at elevated CO<sub>2</sub> concentrations were simply wrong? Not necessarily. Provocatively, Beaufort

*et al.*<sup>1</sup> also discovered one particular coccolithophore morphotype in their modern data set that goes decidedly against the general trend. This strain became much more heavily calcified as CO<sub>2</sub> levels increased and as pH decreased along a sampling transect that ranged from the open ocean to coastal upwelling waters. The strain seems to be genetically similar to the widely distributed coccolithophore morphotype used in the controversial culture studies<sup>7,8</sup>.

This surprising exception to the rule raises new questions. For instance, if there are common strains of coccolithophores that thrive and calcify more at high CO<sub>2</sub> concentrations, why don't they always dominate where seawater CO<sub>2</sub> is elevated? (They obviously don't do this, because if they did, they would have obscured the strong negative correlation between CO<sub>2</sub> and calcification observed by Beaufort *et al.*<sup>1</sup>.) The most likely reason is that many unknown factors also influence the abundance and calcification of coccolithophores. In fact, despite decades of intensive research effort, the environmental controls on coccolithophore growth are probably less well understood than for almost any other major phytoplankton group. What is clear, however, is that the environmental controls involved include many of the same factors that will also change concurrently with CO<sub>2</sub> levels and pH under future global-change scenarios, such as temperature, visible and ultraviolet light intensity, and the availability of nutrients and trace elements<sup>11</sup>.

The next challenge for marine scientists is to try to understand how coccolithophore calcification and ecology will respond to evolutionary selection induced by this complex web of simultaneously changing environmental variables. Only then will we be able to predict what the net outcome will be for the future rain ratio of the ocean, and for the enigmatic phytoplankton group that drives it. ■

**David A. Hutchins** is in the Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA.  
e-mail: dahutch@usc.edu

1. Beaufort, L. *et al.* *Nature* **476**, 80–83 (2011).
2. Paasche, E. *Physiol. Plant. Suppl.* **3**, 1–82 (1964).
3. Ridgwell, A. *et al.* *Biogeosciences* **6**, 2611–2623 (2009).
4. Riebesell, U. *et al.* *Nature* **407**, 364–367 (2000).
5. Zondervan, I., Zeebe, R. E., Rost, B. & Riebesell, U. *Glob. Biogeochem. Cycles* **15**, 507–516 (2001).
6. Feng, Y. *et al.* *Eur. J. Phycol.* **43**, 87–98 (2008).
7. Iglesias-Rodriguez, M. D. *et al.* *Science* **320**, 336–340 (2008).
8. Shi, D., Xu, Y. & Morel, F. M. M. *Biogeosciences* **6**, 1199–1207 (2009).
9. Engel, A. *et al.* *Limnol. Oceanogr.* **50**, 493–507 (2005).
10. Feng, Y. *et al.* *Mar. Ecol. Prog. Ser.* **388**, 13–25 (2009).
11. Boyd, P. W., Strzepek, R., Fu, F. & Hutchins, D. A. *Limnol. Oceanogr.* **55**, 1353–1376 (2010).

# Non-CO<sub>2</sub> greenhouse gases and climate change

S. A. Montzka<sup>1</sup>, E. J. Dlugokencky<sup>1</sup> & J. H. Butler<sup>1</sup>

**Earth's climate is warming as a result of anthropogenic emissions of greenhouse gases, particularly carbon dioxide (CO<sub>2</sub>) from fossil fuel combustion. Anthropogenic emissions of non-CO<sub>2</sub> greenhouse gases, such as methane, nitrous oxide and ozone-depleting substances (largely from sources other than fossil fuels), also contribute significantly to warming. Some non-CO<sub>2</sub> greenhouse gases have much shorter lifetimes than CO<sub>2</sub>, so reducing their emissions offers an additional opportunity to lessen future climate change. Although it is clear that sustainably reducing the warming influence of greenhouse gases will be possible only with substantial cuts in emissions of CO<sub>2</sub>, reducing non-CO<sub>2</sub> greenhouse gas emissions would be a relatively quick way of contributing to this goal.**

Greenhouse gases (GHGs) alter Earth's climate by absorbing energy in the lower atmosphere and re-emitting it. Although anthropogenic emissions of CO<sub>2</sub> contribute most to GHG-induced warming, several other gases, such as methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), ozone-depleting substances (ODSs), hydrofluorocarbons (HFCs), sulphur hexafluoride (SF<sub>6</sub>) and perfluorocarbons (PFCs), also affect climate for decades to millennia after being emitted. Because most anthropogenic emissions of these non-CO<sub>2</sub> GHGs are linked to society's fundamental needs for food and energy, they will continue to increase and further warm the climate unless substantial efforts are undertaken to reduce them worldwide<sup>1–4</sup>. Significant opportunities to mitigate anthropogenic emissions of these gases exist, although some will be easier to exploit than others<sup>2,5–8</sup>. By quantifying emissions associated with different human activities and expressing those emissions on a CO<sub>2</sub>-equivalent basis, climate benefits can be linked with specific mitigation targets.

Here we review human activities and natural processes that emit non-CO<sub>2</sub> GHGs, and explore the time-dependent responses of the warming influence (also called climate forcing or radiative forcing) of these chemicals to emissions reductions. Large cuts in anthropogenic emissions of both long- and short-lived GHGs could stabilize climate forcing fairly rapidly, though the exact impacts of such cuts are uncertain because of feedbacks between climate change and GHG emissions from natural processes. Improving our understanding of these feedbacks and our ability to quantify anthropogenic and natural GHG fluxes on regional and global scales could help in assessing mitigation efforts and ensure their success in diminishing future climate warming.

## Sources and sinks of non-CO<sub>2</sub> gases

### Overall changes in emissions

The influence of an emitted GHG on future climate is estimated from its ability to absorb available infrared radiation and its persistence in the atmosphere. The global warming potential (GWP) of a GHG emission is this climate influence, integrated over time and expressed relative to the climate influence of an equivalent mass of CO<sub>2</sub> emission (Box 1). Multiplying emission amounts by 100-yr GWPs allows estimates of 'CO<sub>2</sub>-equivalent' (CO<sub>2</sub>-eq) emissions. This approach has a number of limitations<sup>9</sup> but does provide a means of comparing time-integrated climate impacts arising from emissions of different GHGs.

During the 1980s, human-derived, non-CO<sub>2</sub> GHG emissions were comparable in sum to CO<sub>2</sub> emissions from fossil fuel combustion

(~20 Gt CO<sub>2</sub>-eq yr<sup>-1</sup> from each; Fig. 1). Since 1990, total emissions of non-CO<sub>2</sub> GHGs have declined to 15 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>, mainly due to reductions in ODSs as agreed to in the Montreal Protocol on Substances that Deplete the Ozone Layer<sup>10,11</sup> (henceforth the Montreal protocol). For a time these reductions were sufficient to offset increases in emissions of other GHGs, but since 2003 continued increases in fossil fuel CO<sub>2</sub> have exceeded those offsets. In 2008, the 15 Gt CO<sub>2</sub>-eq yr<sup>-1</sup> contribution from non-CO<sub>2</sub> GHGs accounted for 30% of all human-related long-lived greenhouse gas (LLGHG) emissions (total anthropogenic CO<sub>2</sub> emissions were ~35 Gt CO<sub>2</sub> yr<sup>-1</sup> in 2008<sup>12</sup>; data available at [http://lgmacweb.env.uea.ac.uk/lequere/co2/carbon\\_budget.htm](http://lgmacweb.env.uea.ac.uk/lequere/co2/carbon_budget.htm); Fig. 1). In this section, we review natural and anthropogenic sources and sinks of non-CO<sub>2</sub> GHGs and the potential for their magnitudes to change with climate.

### Methane

At ~1.8 p.p.m. (parts per million as mole fraction in dry air; refs 13, 14), CH<sub>4</sub> is the most abundant non-CO<sub>2</sub> GHG in the atmosphere today. Its present atmospheric mole fraction is 2.5 times higher than that observed in ice cores dated to AD 1000–1750 and is higher than that observed throughout the existing ice-core record, which spans the past 800,000 yr (refs 15, 16). Methane's atmospheric increase since 1750 implies anthropogenic emissions of 340 ± 50 Tg CH<sub>4</sub> yr<sup>-1</sup> (8.5 ± 1.3 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>; uncertainty, 1 s.d.), or nearly two-thirds of present total emissions, assuming a constant lifetime<sup>1,17,18</sup> (Fig. 2 and Box 1). Agriculture and fossil fuel exploitation together account for about 230 Tg CH<sub>4</sub> yr<sup>-1</sup> (5.8 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>), or two-thirds of all human-derived CH<sub>4</sub> emissions; smaller emissions arise from waste treatment (landfills, manure and sewage) and biomass burning. The contribution of wetlands dominates natural CH<sub>4</sub> emissions (150–180 Tg CH<sub>4</sub> yr<sup>-1</sup>; refs 17, 18) and is highly sensitive to variations in temperature and water table height<sup>19</sup>. Most wetland emissions (~70%) come from the tropics<sup>17,18</sup> and are enhanced during warm, wet periods. Smaller amounts of CH<sub>4</sub> are emitted from oceans and by termites (combined contribution of ≤45 Tg CH<sub>4</sub> yr<sup>-1</sup>; refs 1, 17, 18).

Methane has a shorter steady-state lifetime (~9 yr) than most other LLGHGs<sup>20</sup>, and its loss is primarily the result of atmospheric oxidation by the hydroxyl radical (OH•). This sink links the abundance of CH<sub>4</sub> to the complex atmospheric chemistry that influences OH• concentrations ([OH•]) on a global scale and creates a positive feedback between CH<sub>4</sub>

<sup>1</sup>National Oceanic and Atmospheric Administration, Boulder, Colorado 80305, USA.



## BOX 1

## Estimating GHG emissions

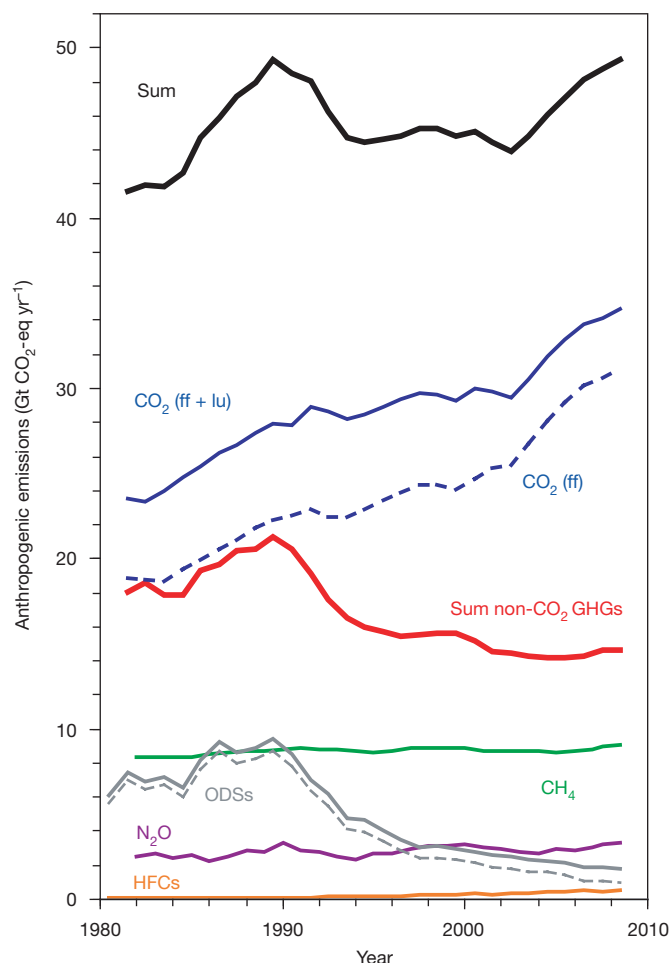
A successful mitigation strategy must include the means to quantify emissions and their changes and include the ability to attribute changes in emissions to specific processes, be they natural or anthropogenic. Emissions are estimated using different methodologies.

- Inventory-based, 'bottom-up' accountings of activities resulting in emissions are multiplied by appropriate emission factors. GHG emissions reported to the United Nations Framework Convention on Climate Change are derived from inventory-based accountings that have a range of accuracies. These inventories are based on an incomplete understanding of emissive processes, have uncertainties that are difficult to quantify and typically are updated infrequently.
- Atmosphere-based studies involve inferring flux magnitudes on local to global scales on the basis of measured atmospheric mixing ratio changes and gradients, their time-dependent variations, and covariations with meteorological parameters or other trace gases. These 'top-down' approaches have historically provided estimates of total global fluxes and can provide independent assessments of global, inventory-based emission magnitudes. Significant discrepancies have been identified<sup>7,61,63,91</sup> between bottom-up and top-down global estimates, highlighting uncertainties in both methods. Recent advances in measurement technologies and high-resolution transport models have made it possible to calculate regional estimates of emissions from top-down analyses<sup>43,44,92–96</sup>. These approaches may provide an independent means to assess or verify regional or national bottom-up emission inventories but at present have substantial uncertainties.
- Process-based approaches provide an understanding of the mechanisms involved in generating GHG emissions and the sensitivities of these mechanisms to changes in climate-related variables such as temperature and precipitation. Knowledge gained from process studies can improve both bottom-up and top-down methodologies.

Emissions are expressed, for example for CH<sub>4</sub>, in units of teragrams of CH<sub>4</sub> per year, where 1 Tg CH<sub>4</sub> is equivalent to  $1 \times 10^{12}$  g CH<sub>4</sub>,  $0.75 \times 10^{12}$  g C and 0.00075 Gt C. Emissions are multiplied by 100-yr GWPs updated in the most recent climate assessment<sup>9</sup> (25 for CH<sub>4</sub>, 298 for N<sub>2</sub>O, 133–14,400 for commonly used ODSs and HFCs, and 7,000–23,000 for PFCs, NF<sub>3</sub> and SF<sub>6</sub>) to give the CO<sub>2</sub>-eq emissions reported here. With this equivalence, an emission of 100 Tg CH<sub>4</sub> corresponds to 2.5 Gt CO<sub>2</sub>-eq. One-hundred-year GWPs adopted by the United Nations Framework Convention on Climate Change are based on previous scientific assessments<sup>47</sup> and would give CO<sub>2</sub>-eq CH<sub>4</sub> emissions 16% smaller, CO<sub>2</sub>-eq N<sub>2</sub>O emissions <0.7% larger and total emissions from other non-CO<sub>2</sub> GHGs ~4% smaller than those estimated here.

and its own abundance; increased CH<sub>4</sub> is expected to decrease [OH\*] and lead to a longer CH<sub>4</sub> lifetime. Reduced [OH\*] would also mean longer lifetimes and enhanced climate influences for HFCs and hydrochlorofluorocarbons (HCFCs), decreased conversion of SO<sub>2</sub> to sulphate aerosol (less cooling) and changes in the cycling of tropospheric ozone, which is a powerful, very short-lived GHG<sup>9,21,22</sup>. Very short-lived substances (for example NO<sub>x</sub> and tropospheric ozone, which have lifetimes of days to months) can still affect climate forcing on longer timescales through their influence on [OH\*] and, therefore, CH<sub>4</sub> (ref. 22).

In the absence of substantial mitigation efforts, both anthropogenic and natural emissions of CH<sub>4</sub> are likely to increase in the twenty-first century as global population increases and climate warms. Inventory-based estimates indicate a substantial increase in human-derived CH<sub>4</sub> emissions in recent years (10% from 2000 to 2005), for example, due to rapid economic growth and increasing demand for food and energy,

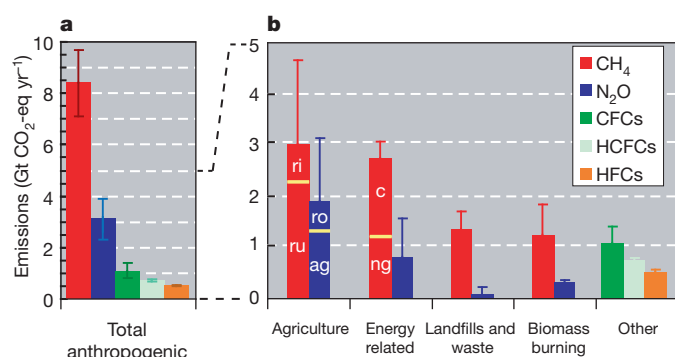


**Figure 1 | Anthropogenic emissions of non-CO<sub>2</sub> GHGs and CO<sub>2</sub>.**

Anthropogenic emissions of non-CO<sub>2</sub> GHGs are calculated as the difference between total emissions derived from observed mixing ratio changes in the global background atmosphere with constant steady-state lifetimes<sup>9,56</sup> and a constant natural background<sup>9,10,13,33,56</sup>. As a result, any variations in natural fluxes are included in the values presented. To reduce the influence of natural variability on our ability to discern decadal changes, emissions of CH<sub>4</sub> and N<sub>2</sub>O were smoothed by averaging over 4-yr periods. By the end of 2008, emissions of CFCs (predominantly from halocarbon banks in recent years; dashed grey line) accounted for just over half of total ODS emissions (CFCs plus HCFCs; solid grey line). Emissions of CO<sub>2</sub> from fossil fuel exploitation (ff) and from fossil fuels plus land use change (ff + lu) are from bottom-up inventories<sup>12</sup>. Weighting factors used here to derive CO<sub>2</sub>-eq emissions (100-yr global warming potentials) are from ref. 9 (Box 1). Emissions too small in 2008 to be visible on this scale include<sup>9,63,64</sup> SF<sub>6</sub> (~0.17 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>), PFCs (~0.1 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and NF<sub>3</sub> (~0.01 Gt CO<sub>2</sub>-eq yr<sup>-1</sup>).

primarily in Asia<sup>23</sup>. Surprisingly, the observed global annual mean atmospheric abundance of CH<sub>4</sub> was nearly constant during this period<sup>13,14</sup> (varying by <0.4% from 1999 to 2006), implying inaccurate inventories, simultaneous offsetting decreases in natural emissions or a comparable increase in [OH\*]. A top-down inverse modelling analysis of global CH<sub>4</sub> observations from 1984 to 2003 also suggests a small increase in total anthropogenic emissions in the early 2000s offset by decreasing natural sources<sup>17</sup>, although these results are not independent of the inventory emissions and their uncertainties used to initialize the model (primarily because of limited available observations).

The sensitivity of natural CH<sub>4</sub> emissions from wetlands to warmer, wetter climates suggests a positive feedback between emissions and climate change that is apparent in ice-core records<sup>15,16</sup>. Although the magnitude of this feedback is uncertain, it could be substantial; modelling of the wetland response to a doubling of CO<sub>2</sub> (corresponding to a global temperature change of 3.4 °C) resulted in an increase in wetland



**Figure 2 | Annual anthropogenic emissions of non-CO<sub>2</sub> GHGs in recent years<sup>1,9,10,17,18,56</sup>.** **a**, Emissions by compound. Uncertainties reflect one standard deviation (s.d.) of estimates<sup>10,17,56</sup>, and for N<sub>2</sub>O are calculated as the difference between total emissions derived from background atmospheric changes<sup>1</sup> and emissions needed to sustain late-Holocene mixing ratios of 255–270 p.p.b. (refs 15, 32), assuming an uncertainty in lifetime of  $\pm 5\%$ . **b**, Emissions by compound and sector (note scale change). Emissions associated with different sources within each main sector are derived from inventory- and atmosphere-based analyses (Box 1). For CH<sub>4</sub>, ruminants (ru,  $2.3 \pm 1.0$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>); rice agriculture (ri,  $0.8 \pm 0.7$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>); natural gas, oil and industry (ng,  $1.6 \pm 0.3$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>); and coal mining (c,  $1.2 \pm 0.2$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) are the main contributors. Values for CH<sub>4</sub> indicate the mean and 1 s.d. of all model estimates in table 7.6 of ref. 1. Emissions of N<sub>2</sub>O are from agriculture directly (ag,  $1.3$  (0.8–2.2) Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and from run-off (ro,  $0.6$  (0.2–0.9) Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) to aquatic ecosystems of excess nitrogen attributable to agriculture<sup>1</sup> (derived as 75% of total nitrogen in run-off<sup>90</sup>). Energy-related N<sub>2</sub>O emissions<sup>1</sup> represent the total from fossil fuel combustion and industrial processes ( $0.33$  (0.1–0.8) Gt CO<sub>2</sub>-eq yr<sup>-1</sup>), atmospheric deposition ( $0.28$  (0.14–0.42) Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and the remaining excess anthropogenic nitrogen reaching aquatic regions from run-off ( $0.2$  (0.06–0.3) Gt CO<sub>2</sub>-eq yr<sup>-1</sup>; 25% of total<sup>90</sup>). Uncertainties quoted here for N<sub>2</sub>O are ranges derived from table 7.7 of ref. 1. Total emissions of CFCs ( $1.0$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>), HCFCs ( $0.7$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and HFCs ( $0.5$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and the indicated uncertainties (1 s.d.) are derived for 2008 from top-down methods<sup>10,56</sup>. Although total biomass burning emissions are included here<sup>9,56</sup> ( $1.3$  Gt CO<sub>2</sub> as CH<sub>4</sub> and  $0.3$  Gt CO<sub>2</sub> as N<sub>2</sub>O), a small fraction of that burning is natural. In **b**, uncertainties shown represent only those associated with the total by compound for each sector.

emissions of CH<sub>4</sub> equal to about one-third of present anthropogenic emissions of CH<sub>4</sub> (ref. 24). In addition, Arctic permafrost contains  $\sim 1,000$  Gt of carbon (Gt C) in its top 3 m, and some of this carbon could be converted to CH<sub>4</sub> by microbes as permafrost melts, depending on the height of the water table<sup>25</sup>. Given the large size of this reservoir, small increases in the release of this carbon as CH<sub>4</sub> as the Arctic warms could substantially augment natural CH<sub>4</sub> emissions. The Arctic Ocean contains an additional 30,000–170,000 Tg CH<sub>4</sub> (20–130 Gt C) in sedimentary clathrates on the relatively shallow continental shelves, although this potential source is unlikely to become important in the short term<sup>26,27</sup>. An additional natural CH<sub>4</sub> source from plants was recently proposed<sup>28</sup> but is probably insignificant<sup>29</sup>.

The near-constant global mixing ratios during 1999–2006 were unexpected, and the next surprise concerning CH<sub>4</sub> was observed in 2007 when its global concentration began increasing again<sup>13,14</sup>. Updated measurements<sup>13</sup> show that this increase is ongoing, at a mean rate during 2007–2010 of  $\sim 6$  p.p.b. yr<sup>-1</sup> (p.p.b., parts per 10<sup>9</sup> as mole fraction in dry air). Assuming no change in the lifetime of CH<sub>4</sub>, these observations correspond to global emissions during 2007–2010 that were  $\sim 17$  Tg CH<sub>4</sub> yr<sup>-1</sup> larger, on average, than in the previous eight years, and they bring up an important question: is this abrupt increase the result of changes in natural processes, human alterations of the natural system or an increase in direct anthropogenic emissions? In the absence of updated inventory estimates, results from global atmospheric observations<sup>13,14</sup> have provided some initial clues as to the causes of the 2007 and 2008 increases. Whereas in 2007 CH<sub>4</sub> mixing ratios increased most rapidly at polar northern latitudes and in the Southern

Hemisphere, in 2008 they increased most rapidly in the tropics, with near zero growth in the Arctic<sup>13</sup>. These observations argue against these mixing ratio increases arising from direct anthropogenic emissions, which would tend to create enhancements in the low- to mid-latitude Northern Hemisphere.

Attributing recent increases to specific CH<sub>4</sub> sources or sinks becomes more definitive with measurements of additional trace gases and related variables. Measurements of CH<sub>4</sub> isotopologues (for example <sup>13</sup>CH<sub>4</sub>) contain information about CH<sub>4</sub> sources because these sources can enrich or deplete the <sup>13</sup>C or <sup>2</sup>H content of atmospheric CH<sub>4</sub>. Methane measured at Alert, Canada, in the summer of 2007 was depleted in <sup>13</sup>C, suggesting an enhanced wetland source in northern high latitudes<sup>13</sup>. Carbon monoxide, ethane and CH<sub>3</sub>Cl are co-emitted with CH<sub>4</sub> from biomass burning, but measurements of these chemicals suggest no significant enhancements during 2007–2008. Measurements of trace gases oxidized by OH<sup>•</sup> (for example methyl chloroform) suggest that [OH<sup>•</sup>] decreases may explain only a portion of the CH<sub>4</sub> enhancements measured in 2007<sup>13,14,30</sup>; changes in [OH<sup>•</sup>] after 2007 have not yet been published. Taken together, the evidence suggests that the renewed increases in atmospheric CH<sub>4</sub> observed during 2007 and 2008 arose primarily from enhanced natural wetland emissions as a result of anomalously high temperatures in the Arctic and greater than average precipitation in the tropics associated with a persistent La Niña<sup>13</sup>. The causes of the continued increases in 2009 and 2010 are not yet clear, but may be related to the strong La Niña that started in early 2010.

### Nitrous oxide

At 322 p.p.b., the atmospheric N<sub>2</sub>O concentration is at present 19% higher than pre-industrial levels<sup>15,31,32</sup> and has increased at a mean rate of  $0.7$  p.p.b. yr<sup>-1</sup> during the past 30 yr (refs 9, 33; data from ref. 33 updated at <http://www.esrl.noaa.gov/gmd/aggi/>). Nitrous oxide concentrations in the pre-industrial atmosphere reflect the balance between production due primarily to biogeochemical cycling of nitrogen and slow stratospheric losses. Humans have augmented nitrogen cycling through the use of inorganic fertilizer, cultivation of nitrogen-fixing crops, and deposition of NO<sub>x</sub> from combustion of fossil fuels<sup>5</sup>. These processes increase the availability of mineralized nitrogen (for example NH<sub>4</sub><sup>+</sup> and NO<sub>3</sub><sup>-</sup>) and lead to N<sub>2</sub>O emissions through microbially mediated nitrification or denitrification in both terrestrial and aquatic ecosystems<sup>5,34,35</sup>. Only a small fraction of applied or deposited anthropogenic reactive nitrogen is ultimately emitted as N<sub>2</sub>O, although this fraction accounts for the largest anthropogenic source of N<sub>2</sub>O and is sensitive to climate, agricultural practices, and the availability of oxygen and nutrients<sup>1,5,34–42</sup>. Other anthropogenic activities such as fossil fuel combustion, industrial processes (for example the production of adipic and nitric acids), biomass burning and waste management release N<sub>2</sub>O directly (Fig. 2).

Anthropogenic N<sub>2</sub>O emissions at present account for  $6.7 \pm 1.7$  Tg N yr<sup>-1</sup> ( $3.1 \pm 0.8$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>; uncertainty, 1 s.d.), or about 40% of all N<sub>2</sub>O emissions, and arise primarily from agricultural activities ( $1.9$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>) and industrial processes including fossil fuel combustion<sup>1</sup> ( $0.8$  Gt CO<sub>2</sub>-eq yr<sup>-1</sup>; Fig. 2). Natural N<sub>2</sub>O sources are predominantly terrestrial ( $\sim 75\%$ ), and these emissions are highest in the tropics<sup>43,44</sup>; natural marine sources typically are strongest in upwelling regions<sup>45</sup>.

Nitrous oxide is removed from the atmosphere much more slowly than CH<sub>4</sub>, given its steady-state lifetime of about 120 yr (refs 46, 47); less than 1% of atmospheric N<sub>2</sub>O is removed annually from the atmosphere, primarily by photolysis and oxidative reactions in the stratosphere. The lifetime of N<sub>2</sub>O may change in the future as it is sensitive to air transport rates through the stratosphere<sup>48</sup> and to stratospheric ozone abundance<sup>49</sup>, but the effects of these processes are poorly quantified or are expected to be small. Nitrous oxide is also destroyed (and produced) by denitrification in nearly anoxic environments and, possibly, as recently suggested, by an uncharacterized process in forest soils under drought<sup>50</sup>; again, however, the influence of these processes on the lifetime of N<sub>2</sub>O is probably small<sup>51</sup>.



How might anthropogenic  $\text{N}_2\text{O}$  emissions change in the future? An inventory analysis<sup>23</sup> suggests increases in  $\text{N}_2\text{O}$  emissions of  $\sim 5\%$  per decade from 1979 to 2005. This is comparable to the increase,  $\sim 7\%$  per decade, inferred for anthropogenic emissions from the observed global increase. This increase is likely to continue in the absence of mitigation efforts<sup>2,4,5</sup>, given that most emissions of  $\text{N}_2\text{O}$  are associated with feeding the world's growing population. Furthermore, enhanced use of fertilizer to cultivate biofuel crops could increase  $\text{N}_2\text{O}$  emissions and offset some of the climate benefits anticipated from biofuels<sup>36–38</sup>. A large fraction of the nitrogen used in agriculture is typically lost to the environment, suggesting that more judicious application of fertilizer could decrease future  $\text{N}_2\text{O}$  emissions without impacting crop yields. Further  $\text{N}_2\text{O}$  mitigation strategies involve increasing nitrogen uptake efficiency by crops, expanding the use of nitrification inhibitors, improving manure management strategies and expanding access to sewage treatment<sup>5,6,8,52</sup>.

Future changes in natural emissions of  $\text{N}_2\text{O}$  are also possible, given ice-core records suggesting that climate change might affect both marine and terrestrial emissions on short and long timescales<sup>15,31,32,41</sup>. Emissions of  $\text{N}_2\text{O}$  from soils in their natural state or soils influenced by humans are sensitive to temperature, hydrology, drought cycles and also land use change<sup>1,5,39–42,50,53,54</sup>. Inversion studies have confirmed the importance of tropical  $\text{N}_2\text{O}$  sources between the equator and  $30^\circ\text{N}$ , and suggest possible emission increases from this region over time<sup>43,44</sup>, although signals are small relative to the present measurement precision and can be dominated by other atmospheric processes such as stratosphere-troposphere exchange<sup>43,44,55</sup>. As is also true for  $\text{CH}_4$ , sensitivities of emissions and loss to natural processes put increased demands on our understanding of regional  $\text{N}_2\text{O}$  budgets in atmosphere-based, top-down inversion studies of emissions, as they make changes in the natural system that are difficult to distinguish from those related to human mitigation efforts.

### Montreal protocol gases and substitutes

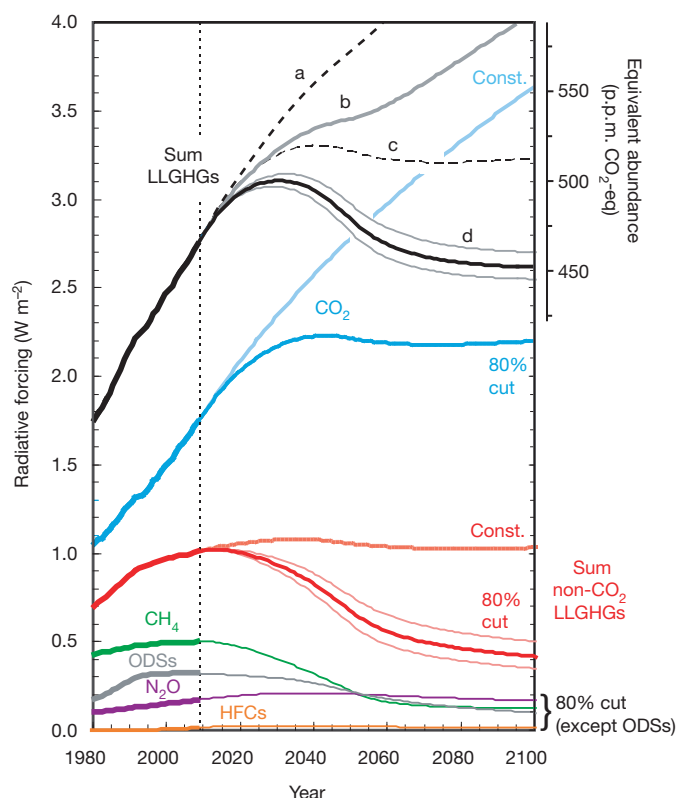
Unlike  $\text{CH}_4$  and  $\text{N}_2\text{O}$ , most ODSs that influence climate have no significant natural sources. Total aggregate emissions of primary ODSs (CFCs, carbon tetrachloride, methyl chloroform, methyl bromide and halons) have decreased from  $\sim 9$  to  $1\text{ Gt CO}_2\text{-eq yr}^{-1}$  since the late 1990s because of adherence to the Montreal protocol (Figs 1 and 2). Global atmospheric concentrations of nearly all these chemicals have declined in recent years as a result<sup>56</sup>. Substitutes for the primary ODSs (that is HCFCs, which are also ODSs controlled by the Montreal protocol, and the non-ozone-depleting HFCs that are included in the Kyoto Protocol) are also potent GHGs. Their global concentrations and emissions have increased in recent years<sup>7,9–11,56</sup>. During 2008, emissions of HCFCs amounted to  $0.7\text{ Gt CO}_2\text{-eq yr}^{-1}$  and emissions of HFCs totalled  $0.5\text{ Gt CO}_2\text{-eq yr}^{-1}$  (Figs 1 and 2).

Although the Montreal protocol's controls on future HCFC production were strengthened recently<sup>10</sup>, future climate impacts of ODSs and HFCs are uncertain for three reasons. First, although HFCs are included in the Kyoto Protocol, their atmospheric burden and, consequently, their radiative forcing could increase substantially owing to their increased use as ODS replacements, particularly in developing countries<sup>10,11</sup>. The future climate impact of HFCs will depend on emission magnitudes and on the extent to which alternatives with shorter lifetimes and lower GWPs are adopted. There is also concern over emissions of HFC-23, a potent GHG created as an unintended by-product during HCFC-22 production. Emission of HFC-23 has the potential to increase in the future in part because only a portion of the HFC-23 is at present destroyed through the Clean Development Mechanism projects<sup>57,58</sup> of the United Nations Framework Convention on Climate Change. Second, halocarbon 'banks', that is, reservoirs of produced but not yet emitted chemicals, are a substantial present source. Most chemicals in banks will ultimately escape to the atmosphere unless destroyed. Although global halocarbon banks of CFCs, HCFCs and HFCs were  $\sim 19\text{ Gt CO}_2\text{-eq}$  in 2005, only a portion is likely to be recoverable<sup>7</sup>. In the United States, for example, about 25% of the 2005 halocarbon bank was classified as 'accessible', that is, in forms

found in fire extinguishing, refrigeration or air conditioning appliances, which are more easily recoverable than, for example, foams<sup>7</sup>. Banks of HCFCs in developing countries and HFCs in developed countries have probably increased in recent years owing to dramatically increased production<sup>7,10</sup>. Third, potential systematic trends in global  $[\text{OH}^\bullet]$  (for example from changes in  $\text{CH}_4$ , carbon monoxide,  $\text{NO}_x$ , tropospheric ozone and water vapour<sup>20,21,59</sup>) would affect HCFC and HFC lifetimes and their time-integrated climate impacts.

### Extremely long-lived gases

A number of chemicals with extremely long lifetimes and small or insignificant natural sources are observed at very low mixing ratios in the



**Figure 3 | Direct radiative forcing derived from observed and projected abundances of LLGHGs<sup>33</sup>.** Projections are from different emissions scenarios: constant future emissions at 2008 levels; 80% cuts in those emissions, phased in linearly between 2009 and 2050 (and constant thereafter); and, for ODSs, emissions based on existing controls supplied by the Montreal protocol<sup>10</sup>. Future forcing from the sum of all LLGHGs (black lines) is shown for constant future emissions (line 'a'), constant future  $\text{CO}_2$  emissions but an 80% decrease in non- $\text{CO}_2$  LLGHG emissions (line 'b'), an 80% decline in  $\text{CO}_2$  emissions but constant future non- $\text{CO}_2$  LLGHG emissions (line 'c') and 80% reductions in emissions of all LLGHGs including  $\text{CO}_2$  (line 'd'). These sums are derived from changes in  $\text{CO}_2$  (blue lines) and the sum of non- $\text{CO}_2$  LLGHGs (red lines). Results for individual non- $\text{CO}_2$  GHGs and groups of them are also shown. Uncertainties shown for selected projections (line 'd' and red line '80% cut') reflect a linear change in  $[\text{OH}^\bullet]$  of  $\pm 15\%$  from 2009 to 2050 (based in part on the sensitivity of  $[\text{OH}^\bullet]$  to  $\text{CH}_4$  changes<sup>21,22</sup>). The small feedback between  $\text{N}_2\text{O}$  abundance and its lifetime, and the small influence of  $\text{CH}_4$  abundance on the  $\text{N}_2\text{O}$  lifetime, were not considered<sup>46</sup>. Future non- $\text{CO}_2$  GHG mixing ratios were calculated with a one-box model, whereas future  $\text{CO}_2$  mixing ratios were calculated using the response function for a pulsed emission decay based on the Bern carbon cycle model<sup>9</sup> (Bern2.5CC). Projections presume constant natural emissions and losses, and thus do not include the potentially significant changes in natural emissions or non- $\text{OH}^\bullet$  losses discussed in the text. For a given emissions scenario, future  $\text{CO}_2$  abundances have additional uncertainties arising from potential changes in uptake by the land and ocean due to the influence of nitrogen availability (see text), climate change, increasing  $\text{CO}_2$  concentration and ocean acidification<sup>1,77–80</sup>. Direct radiative forcings from PFCs ( $0.005\text{ W m}^{-2}$ ),  $\text{SF}_6$  ( $0.004\text{ W m}^{-2}$ ) and  $\text{NF}_3$  ( $0.0001\text{ W m}^{-2}$ ) are too small in 2009 to be visible.

atmosphere today. Changes in climate due to emissions of these chemicals, however, are irreversible on millennial timescales<sup>60</sup>. Examples include SF<sub>6</sub>, SF<sub>5</sub>CF<sub>3</sub>, nitrogen trifluoride (NF<sub>3</sub>) and PFCs (CF<sub>4</sub>, C<sub>2</sub>F<sub>6</sub>, C<sub>3</sub>F<sub>8</sub> and so on). Recent atmospheric changes and inventory estimates<sup>9,56,61–64</sup> indicate fairly small CO<sub>2</sub>-eq emissions of these gases as compared with other LLGHGs (Fig. 1), although emissions for most are increasing, necessitating continued monitoring to ensure that their climate forcings remain small. This is particularly true for gases that could increase substantially with increased electricity use (that is, SF<sub>6</sub>) and for gases not included in existing international protocols (for example NF<sub>3</sub>). Nitrogen trifluoride has received use recently in the electronics industry as a substitute for PFCs owing to climate concerns related to PFCs, and it is largely destroyed during the manufacturing process. Yet the NF<sub>3</sub> that does escape has a larger GWP than the PFCs it replaces<sup>63</sup>.

### Climate forcing from non-CO<sub>2</sub> gases

The direct climate forcing of a trace gas is proportional to its global concentration change since 1750 and how efficiently it absorbs available infrared radiation (its radiative efficiency as watts per square metre per p.p.b.). By 2009, the increase since 1750 in long-lived, non-CO<sub>2</sub> GHGs contributed a direct radiative forcing of 1.0 W m<sup>-2</sup>, or 57% of that from CO<sub>2</sub> (Fig. 3). The concurrent increase in CH<sub>4</sub> accounted for half of this direct climate forcing (0.5 W m<sup>-2</sup>), that in ODSs accounted for about a third (0.32 W m<sup>-2</sup>) and that in N<sub>2</sub>O contributed most of the remainder<sup>33</sup> (0.17 W m<sup>-2</sup>). Smaller forcings (<0.02 W m<sup>-2</sup>) stem from increases in HFCs, PFCs, SF<sub>6</sub> and NF<sub>3</sub> (Fig. 3).

Increases in CO<sub>2</sub> dominated the change in direct radiative forcing from all LLGHGs (0.159 W m<sup>-2</sup>) from mid 2004 to mid 2009; the increase in direct radiative forcing from the sum of all non-CO<sub>2</sub> GHGs was only 20% as large (Fig. 3). Increases in N<sub>2</sub>O (0.012 W m<sup>-2</sup>), CH<sub>4</sub> (0.006 W m<sup>-2</sup>) and HFCs (0.006 W m<sup>-2</sup>) account for most of the change in radiative forcing attributable to non-CO<sub>2</sub> GHGs during this recent five-year period. Increases in direct radiative forcing from SF<sub>6</sub>, PFCs and NF<sub>3</sub> were smaller (<0.001 W m<sup>-2</sup> each), and no net change was observed in direct radiative forcing from ODSs, as increases in HCFCs (0.008 W m<sup>-2</sup>) were offset by comparable decreases in CFCs during this same period.

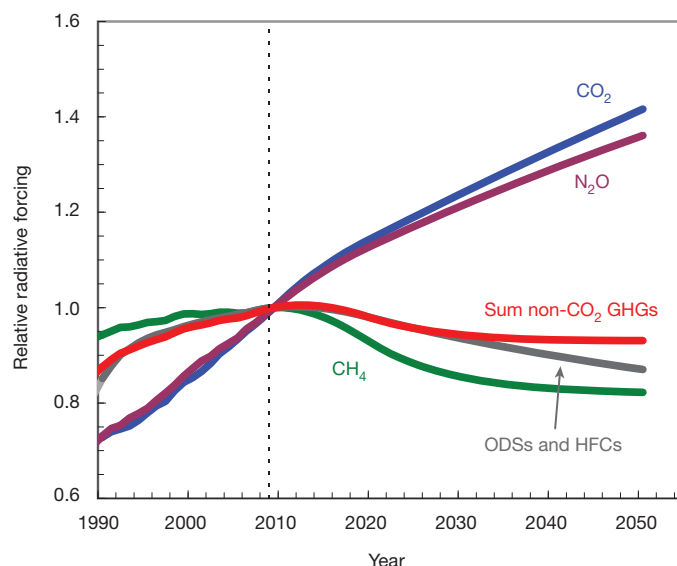
Some non-CO<sub>2</sub> GHGs also influence climate indirectly. For example, increased atmospheric CH<sub>4</sub> enhances stratospheric water vapour and tropospheric ozone, and also affects aerosols<sup>22,65,66</sup>. Including these indirect influences augments the present climate forcing of CH<sub>4</sub> from ~0.5 to 0.8–1.0 W m<sup>-2</sup> (refs 9, 22, 67). Human-derived, short-lived reactive nitrogen (for example NO<sub>x</sub>) affects aerosol abundances, tropospheric ozone, global [OH•] and, therefore, atmospheric lifetimes of CH<sub>4</sub>, HCFCs and HFCs<sup>22</sup>. During the past century, enhanced OH• production from increases in anthropogenic NO<sub>x</sub> may have offset increased OH• losses due to higher concentrations of carbon monoxide and CH<sub>4</sub> (ref. 59). Decreases in stratospheric ozone, mostly from enhanced abundances of ODSs, have caused a climate forcing of  $-0.05 \pm 0.1$  W m<sup>-2</sup> (ref. 9; uncertainty, 90% confidence interval). Including these indirect influences suggests a net climate forcing in 2009 from non-CO<sub>2</sub> GHGs that is 0.2–0.4 W m<sup>-2</sup> higher than when only direct influences are considered.

### Managing climate forcing with non-CO<sub>2</sub> gases

Without substantial mitigation efforts, emissions of non-CO<sub>2</sub> GHGs will probably increase as needs for energy and food increase globally and developing countries continue to modernize<sup>2–5</sup>. An average projected increase of ~50% for emissions of CH<sub>4</sub> and N<sub>2</sub>O by 2050<sup>2,68</sup> (relative to 2000 values) and larger relative increases for HFCs<sup>3,4,10,11,69</sup> correspond to the emission of an additional ~10 Gt CO<sub>2</sub>-eq yr<sup>-1</sup> by 2050. Decreases of ~2 Gt CO<sub>2</sub> yr<sup>-1</sup> are expected for ODS emissions by 2050 with continued adherence to controls set by the Montreal protocol<sup>10</sup>. Considered together, such emissions changes could cause direct climate forcing from all long-lived, non-CO<sub>2</sub> GHGs to reach ~1.5 W m<sup>-2</sup> by 2050 (a 50% increase over their direct radiative forcing today).

To what extent could reductions of non-CO<sub>2</sub> GHG emissions affect climate forcing during the next 50–100 yr? A 25% reduction in anthropogenic emissions, phased in during 2009–2020, illustrates the range of sensitivities GHG concentrations have to emission reductions owing to their different lifetimes. Reducing anthropogenic emissions of all non-CO<sub>2</sub> GHGs in this manner would cause total radiative forcing from these gases to peak during the next decade (Fig. 4). This turnover stems primarily from declines in CH<sub>4</sub> concentration and demonstrates how CH<sub>4</sub> emissions reductions could offset a portion of the increases in radiative forcing from longer-lived GHGs on fairly short timescales. Methane's large influence and rapid response stems from its relatively short lifetime (~9 yr), its near balance of present total emissions and losses, and its large contribution to present climate forcing. This influence depends on the stability of atmospheric [OH•] (Fig. 3), although [OH•] is thought to be buffered against large changes; increases (or decreases) in both CH<sub>4</sub> and NO<sub>x</sub>, for example, have offsetting influences on [OH•] (refs 21, 30, 59). By contrast, a 25% drop in emissions of N<sub>2</sub>O or CO<sub>2</sub> only slightly slows the increase in radiative forcing from each of these long-lived gases, for which present total emissions substantially exceed losses (Fig. 4).

Reversing continued increases in radiative forcing from all LLGHGs will require much larger emissions reductions. An 80% reduction in anthropogenic emissions of CO<sub>2</sub> is required just to stabilize its direct radiative forcing<sup>70</sup> (Fig. 3). Such a large relative decrease is needed because CO<sub>2</sub> in today's atmosphere is long lived<sup>9,71</sup> and because its present total emissions far outweigh losses. Provided that non-CO<sub>2</sub> GHG emissions were also to remain constant, this 80% cut in CO<sub>2</sub> emissions could eventually stabilize climate forcing from all LLGHGs (Fig. 3, line 'c'). By contrast with the time response of CO<sub>2</sub>, an 80% reduction in emissions of all non-CO<sub>2</sub> GHGs by 2050 would cause their radiative forcing to decrease substantially, not just stabilize, mainly owing to the rapid response of CH<sub>4</sub> to emissions decreases. When combined with similar relative cuts in CO<sub>2</sub> emissions, the direct radiative forcing from the sum of all LLGHGs actually peaks and begins decreasing well before 2050 (Fig. 3, line 'd'). But a sustained stabilization of climate forcing cannot be achieved with cuts in only non-CO<sub>2</sub> GHGs (Fig. 3, line 'b').



**Figure 4 | Relative changes in radiative forcing from a 25% cut in GHG emissions.** Radiative forcing expressed as a fraction of the forcing in 2009. Past forcings are based on measured mixing ratios as in Fig. 3. Future forcings are calculated assuming a 25% decline in 2008 emissions, phased in linearly between 2009 and 2020; emissions are kept at 2020 rates thereafter. Projections are made with techniques described in Fig. 3. This illustrative scenario represents an 11% decrease, relative to 1990 levels, in total anthropogenic emissions of GHGs included in the Kyoto Protocol (that is, not ODSs).



A 50–80% cut in CO<sub>2</sub> emissions by 2050 (relative to 1990) has been discussed as a means to limit GHG abundance to <500 p.p.m. CO<sub>2</sub>-eq and mean, global temperature increases to 2–2.4 °C above pre-industrial values<sup>2,70,72</sup>. Such large cuts in CO<sub>2</sub> emissions would probably require substantial changes to future energy production. We considered comparable reductions in non-CO<sub>2</sub> GHG emissions here primarily to demonstrate rough upper limits to and the timescales of climate forcing benefits associated with large cuts in their emissions. Although emissions of non-CO<sub>2</sub> GHGs could be reduced substantially with minimal net cost given present technologies<sup>5,8,68</sup> (40% cuts in CH<sub>4</sub>, 10–15% cuts in N<sub>2</sub>O and ~40% cuts in gases containing fluorine, relative to 2010 values<sup>73</sup>), achieving larger cuts in emissions will depend on cost incentives (for example the price of carbon) and future technological advances<sup>2,68</sup>. Multifaceted modelling analyses that consider abatement costs, mitigation potentials and their potential future evolution demonstrate that a multi-gas approach to stabilizing radiative forcing reduces the required initial cuts in CO<sub>2</sub> emissions and achieves stabilization of climate forcing at a substantially reduced cost relative to mitigating emissions of CO<sub>2</sub> alone<sup>2,4,68</sup>.

These illustrative examples demonstrate the implications of different LLGHG loss timescales for stabilizing and reducing future radiative forcing. But although large cuts in anthropogenic emissions of LLGHGs are clearly needed to stabilize or reduce total direct climate forcing, the exact reductions needed remain somewhat uncertain because natural fluxes of CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O, and a significant fraction of anthropogenic N<sub>2</sub>O emissions, are sensitive to climate. Furthermore, some natural and anthropogenic emissions of N<sub>2</sub>O and CH<sub>4</sub> could increase in response to increasing atmospheric concentrations of CO<sub>2</sub>, although the magnitude of this sensitivity is uncertain<sup>74</sup>.

### Indirect impacts of N<sub>2</sub>O mitigation

Recent studies also suggest that efforts to mitigate some N<sub>2</sub>O emissions may have unintended consequences or offsets affecting other GHGs. In many terrestrial and marine ecosystems, nitrogen availability limits carbon uptake despite additions from human activities that more than double the amount of reactive nitrogen annually added by natural processes to the terrestrial biosphere and augment it by ~25% in the marine environment<sup>41,42,52</sup>. These anthropogenic nitrogen additions have led to increased carbon uptake: approximately 10% of oceanic CO<sub>2</sub> uptake is attributable to the atmospheric transport and deposition of anthropogenic nitrogen, and anthropogenic nitrogen deposited in forests may explain a portion of increases in terrestrial carbon uptake<sup>42,52,75</sup>. Model projections of atmospheric CO<sub>2</sub> in 2100 depend substantially on the magnitude of 'CO<sub>2</sub> fertilization', by which increased CO<sub>2</sub> concentrations lead to enhanced carbon uptake by the biosphere<sup>1</sup>. But simple analyses, coupled models considering nutrient availability and new observational data suggest that future CO<sub>2</sub> fertilization could be reduced partly or entirely by nitrogen limitations<sup>76–81</sup>. These results imply that mitigation efforts to reduce anthropogenic nitrogen additions (and associated N<sub>2</sub>O emissions) from certain activities (particularly fossil fuel combustion and biomass burning) may have offsets related to reduced carbon uptake and increased atmospheric CO<sub>2</sub>. Nitrous oxide mitigation efforts that reduce atmospheric NO<sub>x</sub>, however, may also have indirect climate benefits because they could lead to reduced tropospheric ozone and, potentially, a terrestrial biosphere better able to sequester carbon<sup>82</sup>. Although we cannot at present estimate the overall indirect climate impacts of future N<sub>2</sub>O mitigation efforts, we do know that the magnitude of a climate offset will depend on existing soil nitrogen burdens, loss rates of anthropogenic nitrogen from soils, rates for nitrogen remineralization in a warmer climate and in the presence of increased CO<sub>2</sub> concentrations, the influence of NO<sub>x</sub>-related tropospheric ozone changes on carbon uptake and the adaptability of carbon/nitrogen ratios in organisms to changing nutrient supply ratios<sup>41,52,76,82–87</sup>.

### The way forward

The considerations described in this Review highlight a number of important principles and scientific needs for managing future climate

forcing. First, because non-CO<sub>2</sub> GHGs at present account for about one-third of total CO<sub>2</sub>-eq emissions and 35–45% of total climate forcing from all LLGHGs (range represents direct forcing to the sum of direct and indirect forcing), cuts in their emissions could substantially lessen future climate forcing. Ancillary benefits to mitigating non-CO<sub>2</sub> GHG emissions include reduced costs for climate mitigation relative to CO<sub>2</sub>-only approaches, improvements in air and water quality, reduced acid deposition and decreased eutrophication of aquatic ecosystems<sup>2,4,5,8,68</sup>.

Second, cuts in emissions of the shorter-lived, non-CO<sub>2</sub> GHGs, primarily CH<sub>4</sub>, could cause a rapid decrease in the radiative forcing attributable to these gases. Such a quick response is not possible from cuts in emissions of CO<sub>2</sub> alone. Reducing peak climate forcing and minimizing the time during which it is enhanced could lessen the possibility that the climate irreversibly crosses a tipping point into a new state<sup>67,70</sup>. Although not explicitly considered here owing to their very short lifetimes (<1 yr), other atmospheric constituents such as tropospheric ozone, aerosols and black carbon also affect climate<sup>1,9,22,67,88</sup>. Efforts to reduce their abundances can lead to significant cooling or warming on even shorter timescales than can cuts in CH<sub>4</sub> emissions<sup>67,88</sup>, although both long- and short-lived GHGs influence climate for longer periods than suggested by their lifetimes, owing to time lags associated with heat exchange between the ocean and the atmosphere<sup>60</sup>.

Third, the potential benefits of cuts in non-CO<sub>2</sub> GHG emissions have limits. If, for example, anthropogenic emissions of all non-CO<sub>2</sub> GHGs considered here had been reduced to zero in 2010 and in subsequent years, the decline in direct radiative forcing between 2010 and 2050 would not be sufficient to completely offset continued CO<sub>2</sub> increases at present rates (~1 W m<sup>-2</sup> per 40 yr). Because of these limits, the sustained stabilization of future climate forcing from LLGHGs is only possible with substantial decreases in emissions of CO<sub>2</sub>.

Fourth, the stabilization of climate forcing will be managed more effectively with scientific advances that enhance our understanding of the sensitivities of natural GHG fluxes to climate change and that improve our ability to quantify both natural and anthropogenic GHG fluxes. With such advances, the overall effectiveness of mitigation efforts could be assessed and optimized over time. Improved scientific capabilities could also increase our ability to reduce GHG emissions and may provide a better understanding of the precise cuts in anthropogenic emissions needed to stabilize climate forcing during this century.

The principles outlined above suggest that a research agenda for managing future radiative forcing would include additional process-based studies to elucidate better the mechanisms, sensitivities and linkages likely to affect future non-CO<sub>2</sub> GHG emissions and improve methods of deriving inventory-based emission estimates that form the basis for treaty compliance, carbon-based trading markets and the initialization of top-down estimation techniques. Such an agenda would also include enhancing observational networks and improving inverse modelling techniques to allow more accurate estimates of regional emissions. The 2007–2010 CH<sub>4</sub> anomaly is an example of how existing trace-gas measurement networks and modelling capabilities are only marginally capable of diagnosing specific causes of a substantial change in the balance of its sources and sinks.

Despite these gaps in our present understanding, it is clear that there are substantial climate benefits to managing reductions in emissions of both CO<sub>2</sub> and non-CO<sub>2</sub> LLGHGs. The benefit and success of controls on both long- and short-lived gases have been demonstrated by the Montreal protocol, which resulted in a rapid and sustained decrease in atmospheric concentrations of ozone-depleting chemicals even though the most abundant were long lived<sup>56,89</sup>. Lifetimes of GHGs are similarly varied, and managing emissions reductions in consideration of these timescales could provide a sustained stabilization of climate forcing earlier and at lower peak values than might otherwise be achieved.

1. Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) Ch. 7 (Cambridge Univ. Press, 2007).
2. Fisher, B. S. *et al.* in *Climate Change 2007: Mitigation of Climate Change* (eds Metz, B. *et al.*) Ch. 3 (Cambridge Univ. Press, 2007).

3. Velders, G. J. M., Fahey, D. W., Daniel, J. S., McFarland, M. & Andersen, S. O. The large contribution of projected HFC emissions to future climate forcing. *Proc. Natl Acad. Sci. USA* **106**, 10949–10954 (2010).
4. van Vuuren, D. P., Weyant, J. & de la Chesnaye, F. Multi-gas scenarios to stabilize radiative forcing. *Energy Econ.* **28**, 102–120 (2006).
5. Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**, 889–892 (2008).
6. Vitousek, P. M. *et al.* Nutrient imbalances in agricultural development. *Science* **324**, 1519–1520 (2009).
7. Montzka, S. A., Daniel, J. S., Cohen, J. & Vick, K. in *Trends in Emissions of Ozone-Depleting Substances, Ozone Layer Recovery, and Implications for Ultraviolet Radiation Exposure* (eds Ravishankara, A. R., Kurylo, M. J. & Ennis, C. A.) Ch. 2 (US Department of Commerce, 2008).
8. Townsend, A. R. & Howarth, R. W. Fixing the global nitrogen problem. *Sci. Am.* **302**, 64–71 (2010).
9. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) Ch. 2 (Cambridge Univ. Press, 2007).
10. Daniel, J. S. *et al.* in *Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project—Report No. 52* (ed. Ennis, C. A.) Ch. 5 (World Meteorological Organization, 2011).
11. Velders, G. J. M., Andersen, S. O., Daniel, J. S., Fahey, D. W. & McFarland, M. The importance of the Montreal Protocol in protecting climate. *Proc. Natl Acad. Sci. USA* **104**, 4814–4819 (2007).
- This paper quantified the GWP-weighted reductions in emissions of ozone-depleting substances and was the first to point out the significant climate benefits achieved by the Montreal protocol in relation to those expected from the Kyoto Protocol.**
12. Friedlingstein, P. *et al.* Update on CO<sub>2</sub> emissions. *Nature Geosci.* **3**, 811–812 (2010).
13. Dlugokencky, E. J. *et al.* Observational constraints on recent increases in the atmospheric CH<sub>4</sub> burden. *Geophys. Res. Lett.* **36**, L18803 (2009).
14. Rigby, M. *et al.* Renewed growth of atmospheric methane. *Geophys. Res. Lett.* **35**, L22805 (2008).
15. Spahni, R. *et al.* Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores. *Science* **310**, 1317–1321 (2005).
16. Loulergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH<sub>4</sub> over the past 80,000 years. *Nature* **453**, 383–386 (2008).
17. Bousquet, P. *et al.* Contribution of anthropogenic and natural sources to atmospheric methane variability. *Nature* **443**, 439–443 (2006).
18. Bergamaschi, P. *et al.* Satellite cartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations. *J. Geophys. Res.* **112**, D02304 (2007).
19. Walter, B. P., Heimann, M. & Matthews, E. Modeling modern methane emissions from natural wetlands 1. Model description and results. *J. Geophys. Res.* **106**, 34189–34206 (2001).
20. Prinn, R. G. *et al.* Evidence for variability of atmospheric hydroxyl radicals over the past quarter century. *Geophys. Res. Lett.* **32**, L07809 (2005).
21. Spivakovskiy, C. M. *et al.* Three-dimensional climatological distribution of tropospheric OH: update and evaluation. *J. Geophys. Res.* **105**, 8931–8980 (2000).
22. Shindell, D. T. *et al.* Improved attribution of climate forcing to emissions. *Science* **326**, 716–718 (2009).
23. European Commission Joint Research Centre. EDGARv4.1. *Emission Database for Global Atmospheric Research* (<http://edgar.jrc.ec.europa.eu/overview.php?v=41>) (2010).
24. Shindell, D. T., Walter, B. P. & Faluvegi, G. Impacts of climate change on methane emissions from wetlands. *Geophys. Res. Lett.* **31**, L21202 (2004).
25. McGuire, A. D. *et al.* Sensitivity of the carbon cycle in the Arctic to climate change. *Ecol. Monogr.* **79**, 523–555 (2009).
26. Petrenko, V. V. *et al.* <sup>14</sup>CH<sub>4</sub> measurements in Greenland ice: investigating Last Glacial Termination CH<sub>4</sub> sources. *Science* **324**, 506–508 (2009).
27. Brook, E., Archer, D., Dlugokencky, E., Frolking, S. & Lawrence, D. in *Abrupt Climate Change* Ch. 5 (US Geological Survey, 2008).
28. Keppler, F., Hamilton, J. T. G., Braß, M. & Röckmann, T. Methane emissions from terrestrial plants under aerobic conditions. *Nature* **439**, 187–191 (2006).
29. Nisbet, R. E. *et al.* Emission of methane from plants. *Proc. R. Soc. Lond. B* **276**, 1347–1354 (2009).
30. Montzka, S. A. *et al.* Small interannual variability of global atmospheric hydroxyl. *Science* **331**, 67–69 (2011).
31. Flückiger, J. *et al.* N<sub>2</sub>O and CH<sub>4</sub> variations during the last glacial epoch: insight into global processes. *Glob. Biogeochem. Cycles* **18**, GB1020 (2004).
32. Schilt, A. *et al.* Glacial-interglacial and millennial-scale variations in the atmospheric nitrous oxide concentration during the last 800,000 years. *Quat. Sci. Rev.* **29**, 182–192 (2010).
33. Hofmann, D. J. *et al.* The role of carbon dioxide in climate forcing from 1979 to 2004: introduction of the Annual Greenhouse Gas Index. *Tellus* **58B**, 614–619 (2006).
34. Seitzinger, S. P., Kroeze, C. & Styles, R. V. Global distribution of N<sub>2</sub>O emissions from aquatic systems: natural emissions and anthropogenic effects. *Chemosphere Glob. Chang. Sci.* **2**, 267–279 (2000).
35. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–929 (2008).
36. Crutzen, P. J., Mosier, A. R., Smith, K. A. & Winiwarter, W. N<sub>2</sub>O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmos. Chem. Phys.* **8**, 389–395 (2008).
- This paper proposed that the use of fertilizer to grow certain crops for biofuel could result in N<sub>2</sub>O emissions sufficient to offset the avoided CO<sub>2</sub> from fossil fuel combustion.**
37. Melillo, J. M. *et al.* Indirect emissions from biofuels: how important? *Science* **326**, 1397–1399 (2009).
38. Smeets, E. M. W., Bouwman, L. F., Stehfest, E., van Vuuren, D. P. & Posthuma, A. Contribution of N<sub>2</sub>O to the greenhouse gas balance of first-generation biofuels. *Glob. Change Biol.* **15**, 1–23 (2009).
39. Skiba, U. & Smith, K. A. The control of nitrous oxide emissions from agricultural and natural soils. *Chemosphere Glob. Chang. Sci.* **2**, 379–386 (2000).
40. Keller, M. *et al.* Soil-atmosphere exchange for nitrous oxide, nitric oxide, methane, and carbon dioxide in logged and undisturbed forest in the Tapajós National Forest, Brazil. *Earth Interact.* **9**, 1–28 (2005).
41. Gruber, N. & Galloway, J. N. An Earth-system perspective of the global nitrogen cycle. *Nature* **451**, 293–296 (2008).
42. Duce, R. A. *et al.* Impact of atmospheric anthropogenic nitrogen on the open ocean. *Science* **320**, 893–897 (2009).
- This paper provides a semi-quantitative synthesis of the influence atmospheric anthropogenic nitrogen has on the balance of GHG exchange in the open ocean.**
43. Hirsch, A. I. *et al.* Inverse modeling estimates of the global nitrous oxide surface flux from 1998–2001. *Glob. Biogeochem. Cycles* **20**, GB1008 (2006).
44. Huang, J. *et al.* Estimation of regional emissions of nitrous oxide from 1997 to 2005 using multinetwork measurements, a chemical transport model, and an inverse method. *J. Geophys. Res.* **113**, D17313 (2008).
45. Suntharalingam, P. & Sarmiento, J. L. Factors governing the oceanic nitrous oxide distribution: simulations with an ocean general circulation model. *Glob. Biogeochem. Cycles* **14**, 429–454 (2000).
46. Hsu, J. & Prather, M. J. Global long-lived chemical modes excited in a 3-D chemistry transport model: stratospheric N<sub>2</sub>O, NO<sub>y</sub>, O<sub>3</sub> and CH<sub>4</sub> chemistry. *Geophys. Res. Lett.* **37**, L07805 (2010).
47. Schimel, D. *et al.* in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. & Meira, G.) Ch. 2 (Cambridge Univ. Press, 1996).
48. Butchart, N. & Scaife, A. A. Removal of chlorofluorocarbons by increased mass exchange between the stratosphere and troposphere in a changing climate. *Nature* **410**, 799–802 (2001).
49. Prather, M. J. Time scales in atmospheric chemistry: coupled perturbations to N<sub>2</sub>O, NO<sub>y</sub>, and O<sub>3</sub>. *Science* **279**, 1339–1341 (1998).
50. Goldberg, S. D. & Gebauer, G. Drought turns a Central European Norway spruce forest soil from an N<sub>2</sub>O source to a transient N<sub>2</sub>O sink. *Glob. Change Biol.* **15**, 850–860 (2009).
51. Syakila, A. & Kroeze, C. The global nitrous oxide budget revisited. *Greenh. Gas Meas. Manag.* **1**, 17–26 (2011).
52. Schlesinger, W. H. On the fate of anthropogenic nitrogen. *Proc. Natl Acad. Sci. USA* **106**, 203–208 (2009).
53. Goldberg, S. D., Knorr, K.-H., Blodau, C., Lischke, G. & Gebauer, G. Impact of altering the water table height of an acidic fen on N<sub>2</sub>O and NO fluxes and soils concentrations. *Glob. Change Biol.* **16**, 220–233 (2010).
54. Bouwman, A. F., Boumans, L. J. M. & Batjes, N. H. Modeling global annual N<sub>2</sub>O and NO emissions from fertilized fields. *Glob. Biogeochem. Cycles* **16**, 1080 (2002).
55. Nevison, C. D., Mahowald, N. M., Weiss, R. F. & Prinn, R. G. Interannual and seasonal variability in atmospheric N<sub>2</sub>O. *Glob. Biogeochem. Cycles* **21**, GB3017 (2007).
56. Montzka, S. A. *et al.* in *Scientific Assessment of Ozone Depletion: 2010, Global Ozone Research and Monitoring Project—Report No. 52* (ed. Ennis, C. A.) Ch. 1 (World Meteorological Organization, 2011).
57. Montzka, S. A. *et al.* Recent increases in global HFC-23 emissions. *Geophys. Res. Lett.* **37**, L02808 (2010).
58. Miller, B. R. *et al.* HFC-23 (CHF<sub>3</sub>) emission trend response to HCFC-22 (CHClF<sub>2</sub>) production and recent HFC-23 emission abatement measures. *Atmos. Chem. Phys.* **10**, 7875–7890 (2010).
59. Lelieveld, J., Dentener, F. J., Peters, W. & Krol, M. C. On the role of hydroxyl radicals in the self-cleansing capacity of the troposphere. *Atmos. Chem. Phys.* **4**, 2337–2344 (2004).
60. Solomon, S. *et al.* Persistence of climate changes due to a range of greenhouse gases. *Proc. Natl Acad. Sci. USA* **107**, 18354–18359 (2010).
- This paper points out how the warming associated with a pulsed GHG emission persists longer than the emission lifetime owing to timescales associated with the movement of heat throughout the climate system, particularly into and out of the ocean.**
61. Levin, I. *et al.* The global SF<sub>6</sub> source inferred from long-term high precision atmospheric measurements and its comparison with emission inventories. *Atmos. Chem. Phys.* **10**, 2655–2662 (2010).
62. Sturges, W. T. *et al.* A potent greenhouse gas identified in the atmosphere: SF<sub>5</sub>CF<sub>3</sub>. *Science* **289**, 611–613 (2000).
63. Weiss, R. F., Mühle, J., Salameh, P. & Harth, C. M. Nitrogen trifluoride in the global atmosphere. *Geophys. Res. Lett.* **35**, L20821 (2008).
64. Mühle, J. *et al.* Perfluorocarbons in the global atmosphere: tetrafluoromethane, hexafluoroethane, and octafluoropropane. *Atmos. Chem. Phys.* **10**, 5145–5164 (2010).
65. Shindell, D. T. Climate and ozone response to increased stratospheric water vapor. *Geophys. Res. Lett.* **28**, 1551–1554 (2001).
66. Fiore, A. M. *et al.* Linking ozone pollution and climate change: the case for controlling methane. *Geophys. Res. Lett.* **29**, 1919 (2002).
67. Hansen, J., Sato, M., Ruedy, R., Lacis, A. & Oinas, V. Global warming in the twenty-first century: an alternative scenario. *Proc. Natl Acad. Sci. USA* **97**, 9875–9880 (2000).
- This paper was one of the first to point out that reductions in non-CO<sub>2</sub> GHG emissions could reduce the rate of global warming during the next half-century and lessen the potential for drastic climate change.**



68. van Vuuren, D. P., Eickhout, B., Lucas, P. L. & den Elzen, M. G. J. Long-term multi-gas scenarios to stabilise radiative forcing: exploring costs and benefits within an integrated assessment framework. *Energy J.* **3** (special issue 3), 201–233 (2006).
69. Gschrey, B. & Schwarz, W. *Projections of Global Emissions of Fluorinated Greenhouse Gases in 2050*. Report No. (UBA-FB) 001318 (Federal Environment Agency Germany, 2009).
70. Solomon, S. *et al.* *Climate Stabilization Targets: Emissions, Concentrations, and Impacts over Decades to Millennia* (Natl Acad. Press, 2010).
71. Archer, D. & Brovkin, V. The millennial atmospheric lifetime of anthropogenic CO<sub>2</sub>. *Clim. Change* **90**, 283–297 (2008).
- This paper summarizes the multiple timescales over which CO<sub>2</sub> is removed from the atmosphere, demonstrating how 10–30% of fossil-fuel-derived CO<sub>2</sub> remains in the atmosphere for 1,000–10,000 yr after being emitted.**
72. Towie, N. Scientists issue declaration at Bali. *Nature* doi:10.1038/news.2007.361 (6 December 2007).
73. Lucas, P. L., van Vuuren, D. P., Olivier, J. G. J. & den Elzen, M. G. J. Long-term reduction potential of non-CO<sub>2</sub> greenhouse gases. *Environ. Sci. Policy* **10**, 85–103 (2007).
74. van Groenigen, K. J., Osenberg, C. W. & Hungate, B. A. Increased soil emissions of potent greenhouse gases under elevated CO<sub>2</sub>. *Nature* (in the press).
75. Thomas, R. Q., Canham, C. D., Weathers, K. C. & Goodale, C. L. Increased tree carbon storage in response to nitrogen deposition in the US. *Nature Geosci.* **3**, 13–17 (2010).
76. Arneeth, A. *et al.* Terrestrial biogeochemical feedbacks in the climate system. *Nature Geosci.* **3**, 525–532 (2010).
77. Zaehle, S., Friedlingstein, P. & Friend, A. D. Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophys. Res. Lett.* **37**, L01401 (2010).
78. Thornton, P. E. *et al.* Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model. *Biogeosciences* **6**, 2099–2120 (2009).
79. Hungate, B. A., Dukes, J. S., Shaw, R., Luo, Y. & Field, C. B. Nitrogen and climate change. *Science* **302**, 1512–1513 (2003).
- With a fairly straightforward analysis, this paper outlines how nitrogen availability will limit carbon uptake from CO<sub>2</sub> fertilization in the future.**
80. Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E. & McCurtrie, R. E. CO<sub>2</sub> enhancement of forest productivity constrained by limited nitrogen availability. *Proc. Natl Acad. Sci. USA* **107**, 19368–19373 (2010).
- Much is being learned about how the biosphere will respond to enhanced CO<sub>2</sub> concentrations, and this study, in summarizing results from over a decade of exposing a forest to enhanced CO<sub>2</sub> concentrations, shows how nitrogen limitations limit long-term CO<sub>2</sub> fertilization effects even in the eastern United States.**
81. McCarthy, H. R. *et al.* Re-assessment of plant carbon dynamics at the Duke free-air CO<sub>2</sub> enrichment site: interactions of atmospheric [CO<sub>2</sub>] with nitrogen and water availability over stand development. *New Phytol.* **185**, 514–528 (2010).
82. Sitch, S., Cox, P. M., Collins, W. J. & Huntingford, C. Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature* **448**, 791–794 (2007).
83. Rütting, T., Clough, T. J., Muller, C., Lieffering, M. & Newton, P. C. D. Ten years of elevated atmospheric carbon dioxide alters soil nitrogen transformations in a sheep-grazed pasture. *Glob. Change Biol.* **16**, 2530–2542 (2010).
84. McKinley, D. C., Romero, J. C., Hungate, B. A., Drake, B. G. & Megonigal, J. P. Does deep soil N availability sustain long-term ecosystem responses to elevated CO<sub>2</sub>? *Glob. Change Biol.* **15**, 2035–2048 (2009).
85. Reay, D. S., Dentener, F., Smith, P., Grace, J. & Feely, R. A. Global nitrogen deposition and carbon sinks. *Nature Geosci.* **1**, 430–437 (2008).
86. Janssens, I. A. & Luyssaert, S. Nitrogen's carbon bonus. *Nature Geosci.* **2**, 318–319 (2009).
87. Taylor, P. G. & Townsend, A. R. Stoichiometric control of organic carbon–nitrate relationships from soils to the sea. *Nature* **464**, 1178–1181 (2010).
88. Unger, N. *et al.* Attribution of climate forcing to economic sectors. *Proc. Natl Acad. Sci. USA* **107**, 3382–3387 (2010).
89. Montzka, S. A. *et al.* Present and future trends in the atmospheric burden of ozone-depleting halogens. *Nature* **398**, 690–694 (1999).
90. Seitzinger, S. P. & Kroeze, C. Global distribution of nitrous oxide production and N inputs in freshwater and coastal marine ecosystems. *Glob. Biogeochem. Cycles* **12**, 93–113 (1998).
91. Parrish, D. D. *et al.* Decadal change in carbon monoxide to nitrogen oxide ratio in U.S. vehicular emissions. *J. Geophys. Res.* **107**, 4140 (2002).
92. Bergamaschi, P. *et al.* Inverse modeling of European CH<sub>4</sub> emissions 2001–2006. *J. Geophys. Res.* **115**, D22309 (2010).
93. Stohl, A. *et al.* Hydrochlorofluorocarbon and hydrofluorocarbon emissions in East Asia determined by inverse modeling. *Atmos. Chem. Phys.* **10**, 3545–3560 (2010).
94. Kort, E. A. *et al.* Emissions of CH<sub>4</sub> and N<sub>2</sub>O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations. *Geophys. Res. Lett.* **35**, L18808 (2008).
95. Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G. & Yavin, Y. G. Emissions of greenhouse gases from a North American megacity. *Geophys. Res. Lett.* **36**, L15810 (2009).
96. Peters, W. *et al.* An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc. Natl Acad. Sci. USA* **104**, 18925–18930 (2007).

**Acknowledgements** We acknowledge discussions with P. Tans and J. Daniel, updates to data published in ref. 33 from J. Elkins, G. Dutton and T. Conway, and technical assistance from C. Siso and B. Miller. This work was supported in part by the Atmospheric Composition and Climate Program and the Carbon Cycle Program of NOAA's Climate Program Office.

**Author Contributions** The writing and drafting of figures was led by S.A.M., but all three authors contributed to the writing and to the ideas presented in this review.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.A.M. ([stephen.a.montzka@noaa.gov](mailto:stephen.a.montzka@noaa.gov)).

# Woody cover and hominin environments in the past 6 million years

Thure E. Cerling<sup>1</sup>, Jonathan G. Wynn<sup>2</sup>, Samuel A. Andanje<sup>3</sup>, Michael I. Bird<sup>4</sup>, David Kimutai Korir<sup>3</sup>, Naomi E. Levin<sup>5</sup>, William Mace<sup>1</sup>, Anthony N. Macharia<sup>1</sup>, Jay Quade<sup>6</sup> & Christopher H. Remien<sup>1</sup>

**The role of African savannahs in the evolution of early hominins has been debated for nearly a century. Resolution of this issue has been hindered by difficulty in quantifying the fraction of woody cover in the fossil record. Here we show that the fraction of woody cover in tropical ecosystems can be quantified using stable carbon isotopes in soils. Furthermore, we use fossil soils from hominin sites in the Awash and Omo–Turkana basins in eastern Africa to reconstruct the fraction of woody cover since the Late Miocene epoch (about 7 million years ago).  $^{13}\text{C}/^{12}\text{C}$  ratio data from 1,300 palaeosols at or adjacent to hominin sites dating to at least 6 million years ago show that woody cover was predominantly less than ~40% at most sites. These data point to the prevalence of open environments at the majority of hominin fossil sites in eastern Africa over the past 6 million years.**

There is long-standing debate as to the importance of woody versus herbaceous cover in the evolution of humans over the past 6 million years (Myr)<sup>1–5</sup>. Despite uncertainty as to the nature of the last common ancestor (LCA) that we share with modern chimpanzees<sup>6</sup>, it is widely recognized that the LCA inhabited wooded environments, and that hominin habitats became less wooded after this divergence some 5–8 Myr ago<sup>2,3,5,6</sup>. Woody plants provide shade, shelter and food resources, and as such could have an important role in the evolution of terrestrial mammals, including humans. For example, shade provided by this cover may have influenced thermoregulatory and endurance running adaptations, or nesting and hunting behaviours of early hominins<sup>7–11</sup>.

In broader ecosystem-scale terms, the role of ‘savannah’ ecosystems in hominin evolution remains a subject of debate<sup>1–5</sup>, although it is widely recognized that savannahs may have influenced a variety of hominin adaptations such as bipedalism and dietary adaptations to novel foods<sup>2,3,6–13</sup>. Consideration of the role of savannahs in human evolution began in 1925<sup>14</sup> with the introduction of what is often described as the ‘savannah hypothesis’ and continues today in efforts to reconcile the fossil record of human origins with diverse palaeo-environmental proxies<sup>1–3,5,15</sup>. However, an imprecise and often overly simplistic application of the definition of savannahs hinders progress in the debate over the timing and nature of their role in human evolution. To move past this persistent problem, we develop a relationship between the modern carbon isotope ratio in soils and the amount of woody cover in tropical environments and show that this can be used as a calibration for estimating woody cover of past environments. By using this relationship we can focus on the degree to which habitats were wooded, thereby circumventing any need to apply a functional definition of savannah to past environments where only structure can be inferred.

Stable carbon isotopes in palaeosols are a key means of reconstructing ancient environments, particularly those in the tropics in the past 6 Myr or longer. Woody plants, almost all of which use the  $\text{C}_3$  photosynthetic pathway, would have provided mammals with shade and shelter from the direct sun<sup>16</sup>. Tropical grasses, on the other hand, use the  $\text{C}_4$  photosynthetic pathway<sup>16</sup>. Because these two pathways differ in

their discrimination against  $^{13}\text{CO}_2$  during photosynthesis, the stable carbon isotopic composition of soils can be used as a direct indicator of the fraction of woody cover in tropical ecosystems<sup>17–19</sup>.

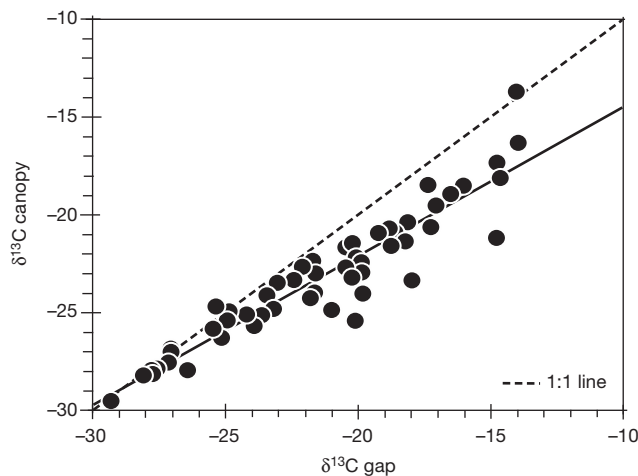
We present a method using stable carbon isotopes to quantify the fraction of woody cover in tropical ecosystems using new data from eastern Africa that builds on earlier observations: the fraction of  $\text{C}_4$  biomass is related to the fraction of woody cover<sup>17,18,20</sup>. We then apply this relationship to well-dated fossil sites in the Awash and Omo–Turkana basins in Ethiopia and Kenya, which contain sedimentary archives that are critical to understanding hominin evolution in the Pliocene and Pleistocene epochs. The in-depth analysis of the relationship between stable carbon isotopes and woody cover in modern soils permits the reconstruction of the woody cover, or shade, available to hominins in eastern Africa during the past 6 Myr.

## Calibrating a ‘palaeo-shade’ proxy

We report results of woody cover measurements and the  $^{13}\text{C}/^{12}\text{C}$  ratio ( $\delta^{13}\text{C}$ ) values of surface horizons from tropical soils for 76 locations in Kenya, Ethiopia, Malaysia, Botswana<sup>20</sup>, Zambia<sup>20</sup>, Australia<sup>17</sup> and Brazil<sup>21</sup> (Supplementary Tables 1 and 2). Sites were selected because of their undisturbed nature (that is, from National Parks or Reserves) or because of minimal agricultural disturbance. ‘Gap’ samples are from soils that are not directly under woody cover and ‘canopy’ samples are from soils directly beneath woody cover. The  $\delta^{13}\text{C}$  values of the gap and the canopy samples are well correlated for the entire data set (Fig. 1; see also Supplementary Information). In closed settings with woody cover of approximately >80%, both the canopy and gap areas show  $\delta^{13}\text{C}$  values characteristic of predominantly  $\text{C}_3$  vegetation<sup>17</sup>. Likewise, open settings, with woody cover approximately <20%, show  $\delta^{13}\text{C}$  values characteristic of predominantly  $\text{C}_4$  vegetation, both in the canopy and gap samples. Thus,  $\delta^{13}\text{C}$  values in these soils reflect the amount of woody cover on the timescales of carbon turnover (~10 yr in the tropics<sup>22</sup>), and  $\text{C}_4$ -rich ecosystems have a different soil  $\delta^{13}\text{C}$  signature than  $\text{C}_3$ -rich ecosystems whether under a tree canopy or not. We invert the relationship in Fig. 2 to reconstruct the fraction of woody cover ( $f_{\text{wc}}$ ) from  $\delta^{13}\text{C}$  values of organic matter and soil carbonate in fossil soils (see methods in Supplementary Information):

<sup>1</sup>University of Utah, Salt Lake City, Utah 84112, USA. <sup>2</sup>University of South Florida, Tampa, Florida 33620, USA. <sup>3</sup>Kenya Wildlife Service, PO Box 40241-00100 Nairobi, Kenya. <sup>4</sup>James Cook University, PO Box 6811, Cairns QLD 4870, Queensland, Australia. <sup>5</sup>Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>6</sup>University of Arizona, Tucson, Arizona 85721, USA.



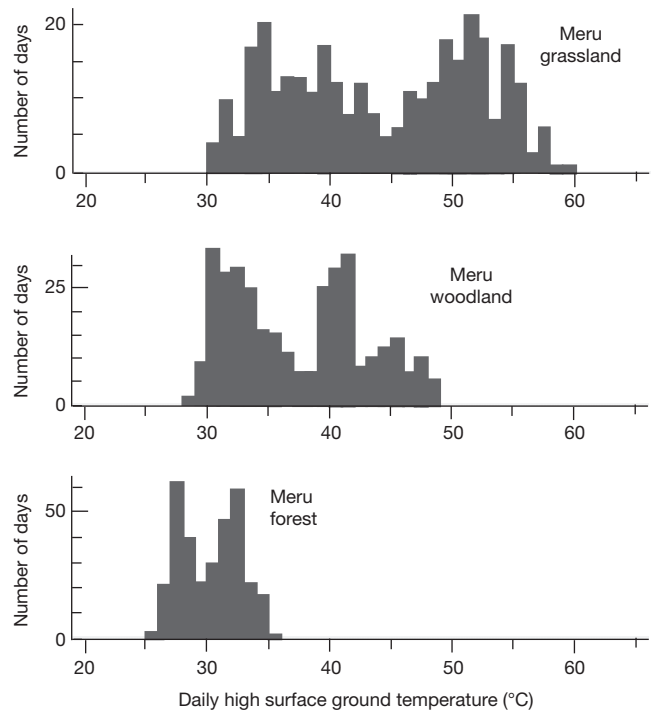


**Figure 1 | Correlation of  $\delta^{13}\text{C}$  between gap and canopy samples for 76 tropical soils used in this study.** Best-fit line is using the major axis regression where  $\delta^{13}\text{C}_{\text{canopy}} = 0.79 \times \delta^{13}\text{C}_{\text{gap}} - 6.4$ ;  $r^2 = 0.89$ .

$$f_{\text{wc}} = \left\{ \sin \left[ -1.06688 - 0.08538 (\delta^{13}\text{C}_{\text{om}}) \right] \right\}^2$$

This relationship is not a simple linear mixing line between  $\text{C}_3$  plants (ca.  $-24$  to  $-35\text{‰}$ ) and  $\text{C}_4$  plants (approximately  $-11\text{‰}$  to  $-14\text{‰}$ ; ref. 16) because  $\text{C}_3$  herbaceous plants may occur in both open and closed areas, and canopy areas may contain both  $\text{C}_3$  trees and both mixed  $\text{C}_3$ – $\text{C}_4$  understory.

Shade provided by woody cover also affects soil temperatures and the proportion of  $\text{C}_3$  and  $\text{C}_4$  biomass in the herbaceous understory. Figure 3 shows maximum daily soil surface temperature for three sites in Meru National Park as calculated from sub-surface soil-temperature loggers for the period June 2009 through to May 2010. Soil ground surface temperatures varied between  $25^\circ\text{C}$  and  $35^\circ\text{C}$  in riparian forest, between  $28^\circ\text{C}$  and  $48^\circ\text{C}$  in woodland, and between  $30^\circ\text{C}$  and  $60^\circ\text{C}$  in grassland. Over a 1-year period, 130 days had daily maximum ground surface temperatures exceeding  $45^\circ\text{C}$  in the grassland (unshaded) environment.  $\text{C}_4$  photosynthesis is an adaption to both high leaf temperature and low atmospheric  $\text{CO}_2$  (ref. 16);  $\text{C}_4$  photosynthesis at higher leaf temperatures is favoured over  $\text{C}_3$  photosynthesis because of increasing photorespiration in  $\text{C}_3$  plants related to increased



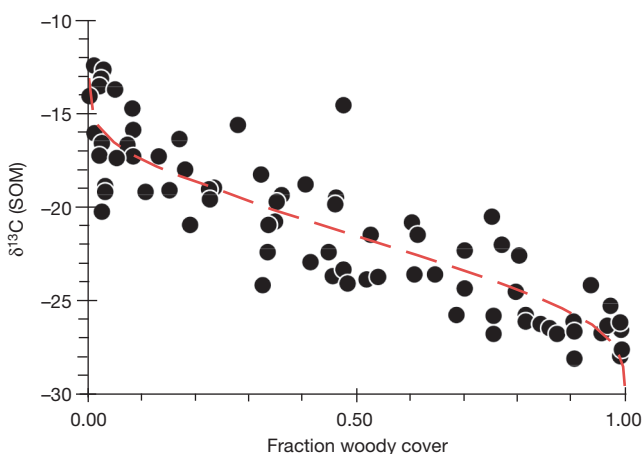
**Figure 3 | Surface soil temperatures from soil temperature profiles.** Calculated maximum daily soil-surface temperatures for a 12-month interval for forest, woodland and grassland sites in the Meru National Park region, Kenya (see Supplementary Tables 1 and 2).

oxygenase activity relative to carboxylase activity. Humidity and soil moisture are also higher in shaded areas. Thus in the areas with higher shade cover due to a woody canopy,  $\text{C}_3$  photosynthesis may be favoured even in the gap areas. Conversely,  $\text{C}_4$  photosynthesis in open areas is favoured by lower daily-integrated shade where there is higher temperature, higher light intensity, with lower relative humidity and soil moisture. Taken together, the abundance of  $\text{C}_3$  biomass in gaps in  $\text{C}_3$ -dominated ecosystems, and the abundance of  $\text{C}_4$  biomass under the canopy in  $\text{C}_4$ -dominated portions of the landscapes, is related to the light intensity and the surface ground temperature, and their effects on photorespiration, humidity and soil moisture.

### A structural classification for palaeo-vegetation

It is useful to define formal terminology of vegetation structural categories like 'forest', 'woodland', 'grassland' before proceeding with our reconstruction of ancient vegetation. Because of the strong relationship between  $\delta^{13}\text{C}$  and  $f_{\text{wc}}$  (Fig. 2), we adopt a vegetation classification system that is based primarily on woody cover (the United Nations Educational, Scientific and Cultural Organization (UNESCO) classification of African vegetation<sup>23</sup>). The principal vegetation types are: (1) forest: a continuous stand of trees at least 10-m tall with interlocking crowns. (2) Woodland/bushland/thicket/shrubland: where woodland is an open stand of trees at least 8-m tall with woody cover  $>40\%$  and a field layer dominated by grasses; bushland is an open stand of bushes usually between 3- and 8-m tall with woody cover  $>40\%$ ; thicket is a closed stand of bushes and climbers usually between 3- and 8-m tall; and shrubland is an open or closed stand of shrubs up to 2-m tall. (3) Wooded grassland: land covered with grasses and other herbs, with woody cover between 10% and 40%. (4) Grassland: land covered with grasses and other herbs, with woody cover  $<10\%$ . (5) Desert: arid landscapes with a sparse cover dominated by sandy, stony or rocky substrate.

Because this classification does not define a boundary between forest and woodland in terms of woody cover, we will consider that 'forest' has  $>80\%$  woody cover based on the requirement for 'interlocking crown



**Figure 2 | Woody cover and soil  $\delta^{13}\text{C}$  for 76 tropical soils used in this study.** Canopy-weighted  $\delta^{13}\text{C}$  values from Supplementary Table 1 have been corrected for the Suess effect (to 1,750)<sup>43,44</sup>, assuming a residence time for carbon in soils to be 10 years<sup>22</sup>. Ordinary linear regression was carried out on arcsine square-root-transformed values of fractional woody cover ( $\arcsin(\sqrt{f_{\text{wc}}})$ )<sup>45</sup>; the dashed line is the OLR function  $\delta^{13}\text{C} = -9.02 \arcsin(\sqrt{f_{\text{wc}}}) - 14.49$  ( $r^2 = 0.77$ ). SOM, soil organic matter.

canopies'. Our reconstruction method cannot distinguish between functionally distinct categories such as woodland, bushland and shrubland, which are defined by the height of woody vegetation.

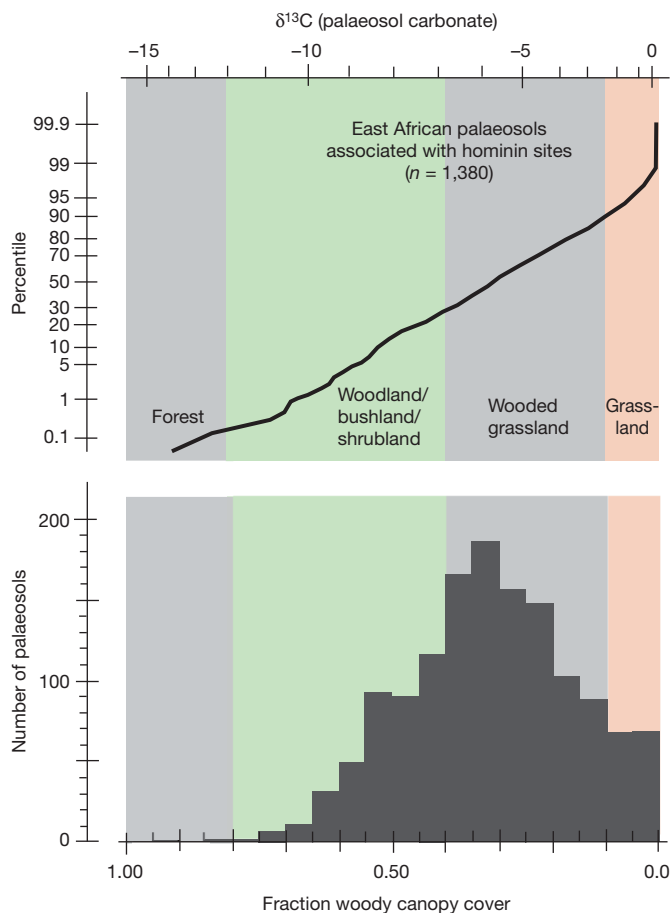
The term savannah suffers from colloquial misuse, and for that reason is not recognized in the UNESCO classification. Still, a modern ecological definition of the term savannah is comprehensive and includes structural, functional and evolutionary aspects<sup>24</sup>. Because our focus is on reconstructing the physiognomic structure of palaeo-vegetation, we use a purely structural definition of savannah: "mixed tree-grass systems characterized by a discontinuous tree canopy in a conspicuous grass layer" (from ref. 24). This, and other common usage of the term<sup>1–3,11,19,25,26</sup>, would include at least 'wooded grasslands' and 'grasslands' in the UNESCO structural categories described above, although woody cover varies significantly within the savannahs (from about 5–80%; refs 24, 27). Rainfall is widely recognized as the primary determinant of woody cover along with tolerance to fire, herbivory and soil fertility<sup>18,24,27,28</sup>. However, variation in these primary determinants, such as between regions or continents or soil types, does not obscure the relationship between  $f_{wc}$  and  $\delta^{13}C$  values of soil organic matter as shown in Fig. 2.

### Application to the fossil record

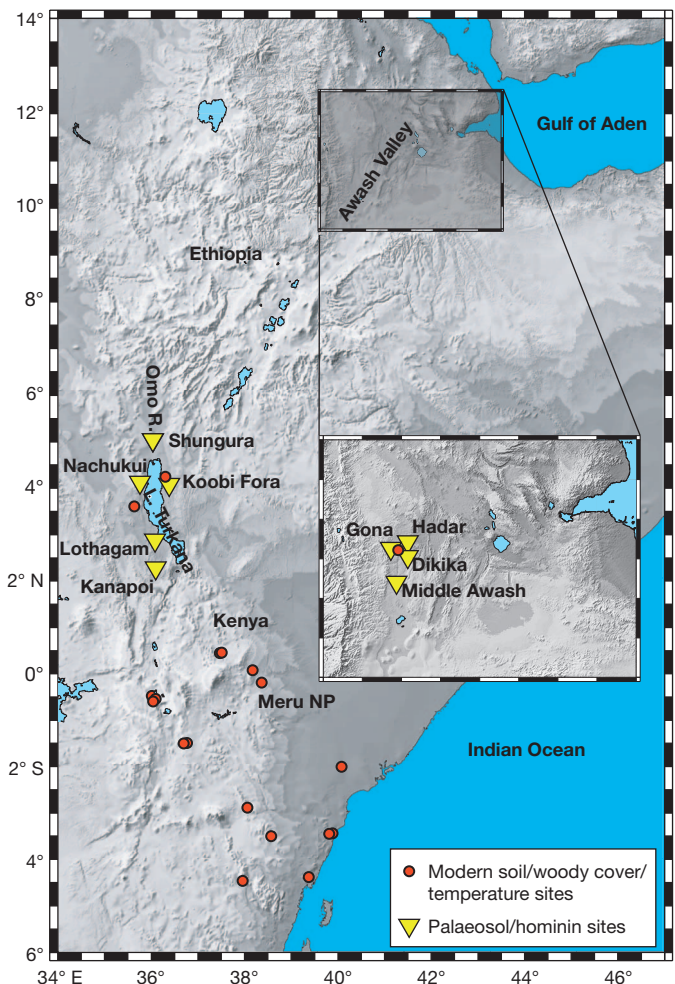
Palaeosols have previously been used to quantify the fraction of  $C_4$  biomass in hominin environments<sup>25,26,29–31</sup>, but the relationship between  $\delta^{13}C$  and woody cover (Fig. 2) provides a means to estimate the fraction of woody cover in past environments. Figure 4 shows a summary of reconstructed woody cover for >1,300 palaeosols associated with

hominin sites in eastern Africa over the past ~6 Myr. More than 70% of these palaeosols reflect woody cover <40%. Less than 1% of the palaeosols are associated with woody cover >70%; therefore, 'closed' forests (>80% woody cover) represent a very small fraction of the environment represented by these palaeosols.

Strata from the Awash Valley and Omo-Turkana Basin span the period from 7.4 Myr ago to the present, and thus bracket the divergence of the LCA of humans and chimpanzees. Much of the hominin fossil record derives from these two basins, which are well dated, and contain abundant palaeosols (Fig. 5 and Supplementary Tables 3–6). We applied our method to long sequences in these two basins to reconstruct changes in the woody cover over this period (Fig. 6 and Supplementary Fig. 1). The record in both basins begins with a period marked by relatively sparse woody cover in the Late Miocene to early Pliocene. At Aramis, where *Ardipithecus ramidus* was found, isotopic compositions are characteristic of grasslands and wooded grasslands<sup>31</sup>, whereas more wooded conditions were present during roughly the same interval for *Ar. ramidus*-bearing sites at Gona. Thus, there is strong evidence for open habitats in the Late Miocene to early Pliocene in the Middle Awash region (approximately 5.7–4.4 Myr ago), as well as at Lothagam (7.4–5.7 Myr ago) in the Omo-Turkana Basin. These generally open habitats are directly associated with the earliest purported members of the Hominini, and this fact should be considered in debate regarding potential mechanisms for novel adaptations of this clade, which include bipedalism, the

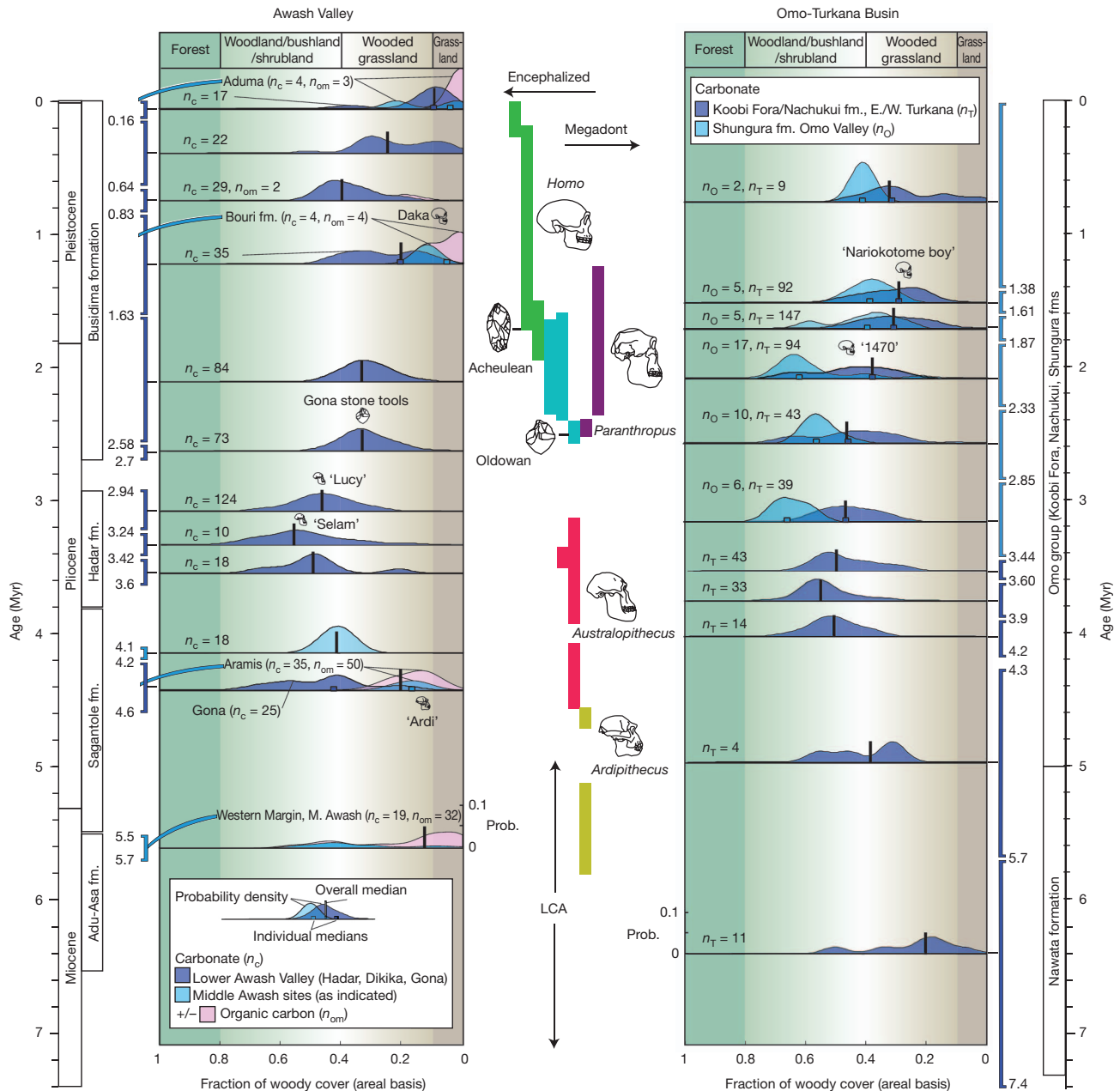


**Figure 4** | Estimated fraction of woody cover based on >1,300 published analyses of palaeosols from eastern African hominin sites from 6 Myr ago to present<sup>25,26,29–31,46–50</sup>. Vegetation classification is from ref. 23. Top: cumulative frequency of palaeosol values related to fraction of woody cover. Bottom: histogram of palaeosol values related to fraction of woody cover.



**Figure 5** | Map with modern soil sites and hominin-fossil bearing localities in the East African Rift System overlain on GTOPO30 digital elevation model. Red circles, modern soil sites; yellow triangles, hominin-fossil-bearing localities.





**Figure 6 | Composite record of palaeosol stable isotopic composition from the Awash Valley, Ethiopia (left) and Omo-Turkana Basin, Kenya and Ethiopia (right).** A hominin phylogram is shown at the centre (adapted from ref. 32). Stable isotope data are presented as normalized probability density functions of predicted woody cover determined for palaeosols in a series of temporal bins defined for each basin. Temporal bins are divided based on marker tephra in each sedimentary basin (see Supplementary Tables 3 and 4). The number of pedogenic carbonate analyses ( $n_c$ ) and of organic matter analyses ( $n_{om}$ ) are indicated for each temporal bin; the median value of woody cover for all data from each temporal bin is shown with a narrow white bar. Data from the Awash Valley include relatively continuous records from co-adjacent research areas at Hadar, Gona and Dikika, along with data from the

advent of megadontia, and diminution of the canine premolar honing complex<sup>5,6,32</sup>.

The relatively open conditions of the Late Miocene to early Pliocene are followed by generally increasing woody cover in the Middle Pliocene (after about 3.6 Myr ago; Fig. 6). Most palaeosols during this time interval in these basins suggest ~40–60% woody cover and limited areas of more open environments. This period of increased woody cover in the Turkana Basin is in part marked by the onset of rapid basin sedimentation in a tectonically driven, basin-wide

lake system, surrounded by generally mesic environments<sup>33</sup>. A Middle Pliocene peak of woody vegetation in the Awash Valley occurs later than that in the Omo-Turkana Basin, but is also associated with a period marked by rapid sedimentation in a tectonically controlled lake that favoured more wooded settings<sup>34</sup>. Thus, the Middle Pliocene of both basins is largely characterized by environments that are more wooded than those in either the Late Miocene (up to about 5.3 Myr ago) or the Pleistocene (~1.8 Myr onwards)—a trend consistent with global biome reconstruction and model-data comparisons that show

lake system, surrounded by generally mesic environments<sup>33</sup>. A Middle Pliocene peak of woody vegetation in the Awash Valley occurs later than that in the Omo-Turkana Basin, but is also associated with a period marked by rapid sedimentation in a tectonically controlled lake that favoured more wooded settings<sup>34</sup>. Thus, the Middle Pliocene of both basins is largely characterized by environments that are more wooded than those in either the Late Miocene (up to about 5.3 Myr ago) or the Pleistocene (~1.8 Myr onwards)—a trend consistent with global biome reconstruction and model-data comparisons that show

expansion of woody vegetation across Africa during the middle Pliocene<sup>35</sup>. These generally more wooded conditions are coincident with the earliest clear evidence of bipedalism, and the earliest widely accepted member of the Hominini, as well as more efficient bipedalism and megadonty of the genus *Australopithecus*<sup>6,32</sup>.

The interval straddling the transition from the Pliocene to Pleistocene (~3.6–1.4 Myr ago) shows the return of open environments such as wooded grasslands and grasslands in these two basins (Fig. 6). The extent of open grasslands peaks during the Pleistocene (~1.8–0.01 Myr ago) and represents the culmination of a longer-term trend towards diminishing woodlands that started in the Late Miocene. These developments, however, were not synchronous across the Omo-Turkana system: more wooded conditions persisted longer in the Omo Valley (into the Late Pliocene, up to about 1.8 Myr ago) than in the lower Turkana Basin. The presence of a large, perennial river draining the highlands of the Omo River basin probably favoured the more wooded environment apparent in these data<sup>36</sup>. Thus, in the environment of the generally open conditions of hot and dry wooded grasslands in the Lower Turkana Basin, areas closer to the axial river system had more interconnected and dense woody cover, nourished by more abundant groundwater and surface water. After 1.9 Myr ago, environments with >50% woody cover nearly disappear from the stratigraphic record in the basin, including areas previously characterized by more wooded conditions, such as the Omo River valley. Woody cover in the Awash Valley also decreases after about 2.9 Myr ago, which is coincident with a shift from lacustrine to fluvial deposition<sup>36</sup>. This overall trend in both basins may be attributable to coeval global and regional climate change<sup>4,15,37–40</sup>, although a pronounced reorganization of fault blocks within the Ethiopian Rift System, and shift from lacustrine to fluvial deposition in Hadar, Gona and Dikika areas<sup>34</sup>, may also have contributed to the decrease in the woody cover after ~2.9 Myr ago. A lack of abundant woody cover persists throughout the Awash Valley during the Pleistocene, with several important sites marked by environments with <10% woody cover. This expansion of grasslands across the Pliocene–Pleistocene transition has been linked to global climate change and major developments in the hominin clade<sup>3,4,15,40</sup>, such as the more obligate bipedalism of the genus *Homo*, increase in encephalization, and reduction in tooth and gut proportions<sup>32</sup>.

## Discussion

Palaeosol stable carbon isotope data from the Omo-Turkana and Awash River basins provide a strong basis for reconstructing the ecological context of hominin evolution, reinforcing debate over mechanisms of co-evolutionary ecology. The evolution of early hominins within African savannahs has been a core principle of the savannah hypothesis proposed by Raymond Dart to explain the origin of the first Pliocene hominin found in Africa (ref. 14; see also refs 1, 2). Since then, debate has evolved from an initial, simplistic 'habitat specific' savannah hypothesis, through more sophisticated notions of co-evolution of hominins and their environment<sup>3,4</sup>, to more recent total rejection of the savannahs as a causal mechanism for the divergence from the LCA<sup>5</sup>. Using the long-standing recognition that tropical savannahs are predominantly mixed tree–C<sub>4</sub> grass systems<sup>24</sup>, the stable isotope evidence from palaeosols discussed in this paper shows that tropical savannahs have been present in East African hominin sites for at least 6 Myr; within the range of savannah environments, the open grassland savannah is much less abundantly represented than the wooded grassland savannah. Thus, within the broadly defined savannah biome, variation in woody cover may have been significant for hominin morphological and behavioural adaptations.

The evidence for the persistence of open vegetation since our divergence from the LCA supports results from a study that used clumped isotope abundance from some of the same soil carbonates to reconstruct soil palaeotemperature<sup>41</sup>, which found evidence for elevated soil temperatures, and related this to the persistence of open vegetation

throughout this basin during the past 4 Myr. Open areas can have daily high surface ground temperatures—up to 25 °C higher than in nearby well-shaded riparian forests or woodlands (Fig. 3). This is because in shaded areas, incident solar radiation is partially balanced by latent and sensible heat transfer from the surface area of trees, whereas in open areas of dry soil, sensible heat flux to the soil predominates. Early hominins would have been affected by this uneven distribution of heat, and this may have influenced physiological and behavioural adaptations that occurred since divergence from the LCA<sup>2,7–13</sup>.

Our observations of the environment of some of the earliest hominins do not contradict the longstanding hypothesis that savannahs in Africa may have had a role in the development of bipedal locomotion, or other key defining characteristics of hominins post-dating the LCA. If either species of *Ardipithecus* (*Ar. ramidus* or *Ar. kadabba*) is validated as the “long-sought potential root species for the Hominidae”<sup>342</sup> then the soil carbonate data now make it clear that both species were surrounded by more open environments than *Australopithecus*, which was more efficiently bipedal and occurred in more wooded environments of both the Omo-Turkana Basin and the Awash Valley (Fig. 6). Thus, the combined results from two of the most significant hominid-bearing regions in eastern Africa leave the savannah hypothesis as a viable scenario for explaining the context of earliest bipedalism, as well as potentially later evolutionary innovations within the hominin clade.

## METHODS SUMMARY

Soils in Kenya and Ethiopia were selected because of their undisturbed nature (National Parks or Reserves) or because of minimal agricultural disturbance. Soil samples (0–5 cm depth) were collected following a protocol<sup>17,20</sup> whereby multiple soils from canopy and gap areas are collected in each site over an area comprising several hectares. Canopy cover was estimated using several methods as described in the Supplementary Information.

Stable carbon isotopes were measured on the organic fraction in modern soils remaining after removal of carbonates on an isotope ratio mass spectrometer operating in continuous-flow mode after combustion at 1,600 °C in an elemental analyser. Results are reported in the standard per mil (‰) notation:  $\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1000$  using the isotope standard Vienna Pee Dee Belemnite. Overall  $\delta^{13}\text{C}$  values of the sites were estimated using the canopy-weighted values for gap and canopy samples<sup>20</sup>.  $\delta^{13}\text{C}$  values of modern soils were corrected for the change in the  $\delta^{13}\text{C}$  of the atmosphere<sup>43,44</sup>, assuming a 10-year residence time for carbon in the surface horizon of tropical soils<sup>22</sup>. Palaeosol carbonates were analysed on CO<sub>2</sub> produced by reaction with H<sub>3</sub>PO<sub>4</sub> at 90 °C.

The several sections of the Supplementary Methods provide additional details of the methods used.

Received 27 January; accepted 15 June 2011.

- Jolly, C. J. *Early Hominids of Africa* (Duckworth, 1978).
- Klein, R. G. *The Human Career: Human Biological and Cultural Origins* (Univ. Chicago Press, 1999).
- Potts, R. Environmental hypotheses of hominin evolution. *Yearb. Phys. Anthropol.* **107**, 93–136 (1998).
- Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H. in *Paleoclimate and Evolution with Emphasis on Human Origins* (Yale Univ. Press, 1995).
- White, T. D. *et al.* *Ardipithecus ramidus* and the Paleobiology of Early Hominids. *Science* **326**, 75–86 (2009).
- Wood, B. & Harrison, T. The evolutionary context of the first hominins. *Nature* **470**, 347–352 (2011).
- Bramble, D. M. & Lieberman, D. E. Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352 (2004).
- Pickering, T. R. & Bunn, H. T. The endurance running hypothesis and hunting and scavenging in savanna-woodlands. *J. Hum. Evol.* **53**, 434–438 (2007).
- Rogers, M. J., Harris, J. W. K. & Feibel, C. S. Changing patterns of land use by Plio-Pleistocene hominids in the Lake Turkana Basin. *J. Hum. Evol.* **27**, 139–158 (1994).
- Sept, J. Shadows on a changing landscape: comparing nesting patterns of hominids and chimpanzees since their last common ancestor. *Am. J. Primatol.* **46**, 85–101 (1998).
- Wheeler, P. E. The thermoregulatory advantages of heat storage and shade seeking behaviour to hominids foraging in equatorial savanna environments. *J. Hum. Evol.* **26**, 339–350 (1994).
- Richmond, B. G., Begun, D. R. & Strait, D. S. The origin of human bipedalism: the knuckle-walking hypothesis revisited. *Yearb. Phys. Anthropol.* **116**, 70–105 (2001).



13. Teaford, M. F. & Ungar, P. S. Diet and the evolution of the earliest human ancestors. *Proc. Natl Acad. Sci. USA* **97**, 13506–13511 (2000).
14. Dart, R. A. *Australopithecus africanus*: the man-ape of South Africa. *Nature* **115**, 195–199 (1925).
15. Behrensmeyer, A. K. Climate change and human evolution. *Science* **311**, 476–478 (2006).
16. Pearcy, R. W. & Ehleringer, J. R. Comparative ecophysiology of C<sub>3</sub> and C<sub>4</sub> plants. *Plant Cell Environ.* **7**, 1–13 (1984).
17. Wynn, J. G. & Bird, M. I. Environmental controls on the stable carbon isotopic composition of soil organic carbon: implications for modeling the distribution of C<sub>3</sub> and C<sub>4</sub> plants, Australia. *Tellus B Chem. Phys. Meteorol.* **60**, 604–621 (2008).
18. Lloyd, J. *et al.* Contributions of woody and herbaceous vegetation to tropical savanna ecosystem productivity: a quasi-global estimate. *Tree Physiol.* **28**, 451–468 (2008).
19. Cerling, T. E. *et al.* Comment on the Paleoenvironment of *Ardipithecus ramidus*. *Science* **328**, doi:10.1126/science.1185274. (2010).
20. Bird, M. I., Veenendaal, E. M. & Lloyd, J. Soil carbon inventories and  $\delta^{13}\text{C}$  along a moisture gradient in Botswana. *Glob. Change Biol.* **10**, 342–349 (2004).
21. Ometto, J. P. *et al.* The stable carbon and nitrogen isotopic composition of vegetation in tropical forests of the Amazon Basin, Brazil. *Biogeochemistry* **79**, 251–274 (2006).
22. Bird, M. I., Chivas, A. R. & Head, J. A latitudinal gradient in carbon turnover times in forest soils. *Nature* **381**, 143–146 (1996).
23. White, F. *The Vegetation of Africa* Vol. 20 (United Nations Scientific and Cultural Organization, 1983).
24. Ratnam, J. *et al.* When is a 'forest' a savanna, and when does it matter. *Glob. Ecol. Biogeogr.* doi:10.1111/j.1466-8238.2010.00634.x (2011).
25. Cerling, T. E. Development of grasslands and savannas in East Africa during the Neogene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 241–247 (1992).
26. Wynn, J. G. Influence of Plio-Pleistocene aridification on human evolution from paleosols of the Turkana Basin, Kenya. *Am. J. Phys. Anthropol.* **123**, 106–118 (2004).
27. Sankaran, M. *et al.* Determinants of woody cover in African savannas. *Nature* **438**, 846–849 (2005).
28. Good, S. P. & Caylor, K. K. Climatological determinants of woody cover in Africa. *Proc. Natl Acad. Sci. USA* **108**, 4902–4907 (2011).
29. Plummer, T. W. *et al.* Oldest evidence of toolmaking hominins in a grassland-dominated ecosystem. *PLoS ONE* **4**, e7199 (2009).
30. Sikes, N. E. & Ashley, G. M. Stable isotopes of pedogenic carbonates as indicators of paleoecology in the Plio-Pleistocene (upper Bed I), western margin of the Olduvai Basin, Tanzania. *J. Hum. Evol.* **53**, 574–594 (2007).
31. WoldeGabriel, G. *et al.* The geological, isotopic, botanical, invertebrate and lower vertebrate surroundings of *Ardipithecus ramidus*. *Science* **326**, 65 (2009).
32. Wood, B. A. & Loneragan, N. The hominin fossil record: taxa, grades and clades. *J. Anat.* **212**, 354–376 (2008).
33. Feibel, C. S., Harris, J. M. & Brown, F. H. in *Koobi Fora Research Project* Vol. 3 (ed. Harris, J. M. Harris) 321–346 (Clarendon, 1991).
34. Quade, J. & Wynn, J. G. in *Geological Society of America Special Publications* Vol. 446 (Geological Society of America, 2008).
35. Salzmann, U., Haywood, A. M., Lunt, D. J., Valdes, P. J. & Hill, D. J. A new global biome reconstruction and data-model comparison for the Middle Pliocene. *Glob. Ecol. Biogeogr.* **17**, 432–447 (2008).
36. Levin, N. E., Brown, F. H., Behrensmeyer, A. K., Bobe, R. & Cerling, T. E. Paleosol carbonates from the Omo Group: Isotopic records of local and regional environmental change in East Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* doi:10.1016/j.palaeo.2011.04.026 (2011).
37. Cane, M. A. & Molnar, P. Closing of the Indonesian seaway as a precursor to east African aridification around 3–4 million years ago. *Nature* **411**, 157–162 (2001).
38. deMenocal, P. B. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth Planet. Sci. Lett.* **220**, 3–24 (2004).
39. deMenocal, P. B. Climate and human evolution. *Science* **331**, 540–542 (2011).
40. Feakins, S. J., deMenocal, P. B. & Eglinton, T. I. Biomarker records of late Neogene changes in northeast African vegetation. *Geology* **33**, 977–980 (2005).
41. Passey, B. H., Levin, N. E., Cerling, T. E., Brown, F. H. & Eiler, J. M. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. *Proc. Natl Acad. Sci. USA* **107**, 11245–11249 (2010).
42. Gibbons, A. A new kind of ancestor: *Ardipithecus ramidus*. *Science* **326**, 36–40 (2009).
43. Francey, R. J. *et al.* A 1000-year high precision record of  $\delta^{13}\text{C}$  in atmospheric CO<sub>2</sub>. *Tellus B Chem. Phys. Meteorol.* **51**, 170–193 (1999).
44. Keeling, R. F., Piper, S. C., Bollenbacher, A. F. & Walker, S. J. Monthly atmospheric  $^{13}\text{C}/^{12}\text{C}$  isotopic ratios for 11 SIO stations. In *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, 2010).
45. Zar, J. H. *Biostatistical Analysis* 2nd edn (Prentice-Hall, 1984).
46. Aronson, J. L., Hailemichael, M. & Savin, S. M. Hominid environments at Hadar from paleosol studies in a framework of Ethiopian climate change. *J. Hum. Evol.* **55**, 532–550 (2008).
47. Levin, N. E., Quade, J., Simpson, S. W., Semaw, S. & Rogers, M. J. Isotopic evidence for Plio-Pleistocene environmental change at Gona, Ethiopia. *Earth Planet. Sci. Lett.* **219**, 93–110 (2004).
48. Quinn, R. L., Lepre, C. J., Wright, J. D. & Feibel, C. S. Paleogeographic variations of pedogenic carbonate  $\delta^{13}\text{C}$  values from Koobi Fora, Kenya: implications for floral compositions of Plio-Pleistocene hominin environments. *J. Hum. Evol.* **53**, 560–573 (2007).
49. White, T. D. *et al.* Asa Issie, Aramis, and the origin of *Australopithecus*. *Nature* **440**, 883–889 (2006).
50. Wynn, J. G. Paleosols, stable carbon isotopes and paleoenvironmental interpretation of Kanapoi, Northern Kenya. *J. Hum. Evol.* **39**, 411–432 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the governments of Kenya and Ethiopia for permission to conduct this research, and F. H. Brown for discussions. The authors thank Kenya Wildlife Service and members of the Dikika and Gona Research Projects for support in the field, and Z. Bedaso for aid with analyses. This research was supported by funding from the LSB Leakey Foundation and NSF grants BCS 0621542, EAR-0617010, EAR-0937819 and BCS-0321893.

**Author Contributions** S.A.A., M.I.B., T.E.C., D.K.K., N.E.L., W.M., J.Q. and C.H.R. designed the modern soil surveys. T.E.C., M.I.B., A.N.M., W.M. and J.G.W. evaluated the amount of woody cover. C.H.R. and W.M. analysed the soil temperature data. M.I.B., N.E.L., A.N.M. and J.G.W. analysed modern soils. N.E.L., J.Q. and J.G.W. contributed new palaeosol data. T.E.C. and J.G.W. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.E.C. ([thure.cerling@utah.edu](mailto:thure.cerling@utah.edu)).

# On the growth and form of the gut

Thierry Savin<sup>1†\*</sup>, Natasza A. Kurpios<sup>2†\*</sup>, Amy E. Shyer<sup>2\*</sup>, Patricia Florescu<sup>1</sup>, Haiyi Liang<sup>1†</sup>, L. Mahadevan<sup>1,3,4,5,6,7</sup> & Clifford J. Tabin<sup>2</sup>

**The developing vertebrate gut tube forms a reproducible looped pattern as it grows into the body cavity. Here we use developmental experiments to eliminate alternative models and show that gut looping morphogenesis is driven by the homogeneous and isotropic forces that arise from the relative growth between the gut tube and the anchoring dorsal mesenteric sheet, tissues that grow at different rates. A simple physical mimic, using a differentially strained composite of a pliable rubber tube and a soft latex sheet is consistent with this mechanism and produces similar patterns. We devise a mathematical theory and a computational model for the number, size and shape of intestinal loops based solely on the measurable geometry, elasticity and relative growth of the tissues. The predictions of our theory are quantitatively consistent with observations of intestinal loops at different stages of development in the chick embryo. Our model also accounts for the qualitative and quantitative variation in the distinct gut looping patterns seen in a variety of species including quail, finch and mouse, illuminating how the simple macroscopic mechanics of differential growth drives the morphology of the developing gut.**

Understanding morphogenesis, the origin of shape in anatomical structures, organs and organisms, has always been a central goal of developmental biology. Historically, the subject focused on the morphology and dynamics of embryonic growth<sup>1</sup>, with many analogies to observable physical phenomena. This metaphoric approach to biological shape is epitomized in D'Arcy Thompson's *On Growth and Form*<sup>2</sup>, with its focus on a mathematical and physical approach to the subject, emphasizing the role of differential growth in determining form. However, with the modern revolution in molecular biology, the field focused on a framework built around gene regulation, signalling molecules and transcription factors. This led to much insight into the logic of the developmental networks controlling processes as diverse as the patterning of the limb skeleton<sup>3</sup> and the branching morphogenesis of the lung<sup>4</sup>. More recently, however, there has been a renewed appreciation of the fact that to understand morphogenesis in three dimensions, it is necessary to combine molecular insights (genes and morphogens) with knowledge of physical processes (transport, deformation and flow) generated by growing tissues.

In this context, there has been only recent limited exploration of the role of tissue-scale mechanical forces in organogenesis<sup>5–10</sup>. Such large-scale forces can become important when the shape of an organ is remodelled after its initial structure has been formed. An important example of this hierarchy is the looping morphogenesis of the gut<sup>11</sup>. The midgut forms as a simple linear tube of circular cross-section running down the midline of the embryo, and grows at a greater rate than the surrounding tissue, eventually becoming significantly longer than the trunk. As the size of the developing mid- and hindgut exceeds the capacity of the embryonic body cavity, a primary loop is forced ventrally into the umbilicus (in mammals) or yolk stalk (in birds). This loop first rotates anticlockwise by 90° and then by another 180° during the subsequent retraction into the body cavity. Eventually, the rostral half of the loop forms the midgut (small intestine) and the caudal half forms the upper half of the hindgut (the ascending colon).

The chirality of this gut rotation is directed by left–right asymmetries in cellular architecture that arise within the dorsal mesentery<sup>12–14</sup>, an initially thick and short structure along the dorsal–ventral axis through which the gut tube is attached to the abdominal wall. This leads the mesentery to tilt the gut tube leftwards with a resulting anticlockwise corkscrewing of the gut as it herniates<sup>12,13</sup>. However, the gut rotation is insufficient to pack the entire small intestine into the body cavity, and additional loops are formed as the intestine bends and twists even as it elongates. Once the gut attains its final form, which is highly stereotypical in a given species, the loops retract into the body cavity. During further growth of the juvenile, no additional loops are formed<sup>15</sup>, as they are tacked down by fascia, which restrict movement and additional morphogenesis without inhibiting globally uniform growth.

## Relative growth between gut and attached mesentery drives looping

Throughout development, the gut tube remains attached to the body wall along its entire length by the dorsal mesentery, and is fixed at both its rostral and its caudal ends to the mouth and anus, respectively, resulting in the preservation of its connectivity and chirality during growth. The resulting number, shape and size of loops are also conserved in any given species, as shown in Fig. 1a for the chick at embryonic day 16 (E16).

In principle, this regularity of looping could result from either the intrinsic properties of the gut tube and mesentery or from external spatial packing constraints. However, surgical dissection of the gut and mesentery from the rest of the embryonic tissues shows that all the loops remain intact and identical to their *in ovo* structure at various stages of development (Fig. 1a), ruling out any role for body-cavity-induced constraints. Another possible mechanism for the reproducible looping is an increased asymmetric proliferation of cells in the gut tube at the locations of the bends. To test this, we counted the numbers of mitotic cells in the entire midgut section

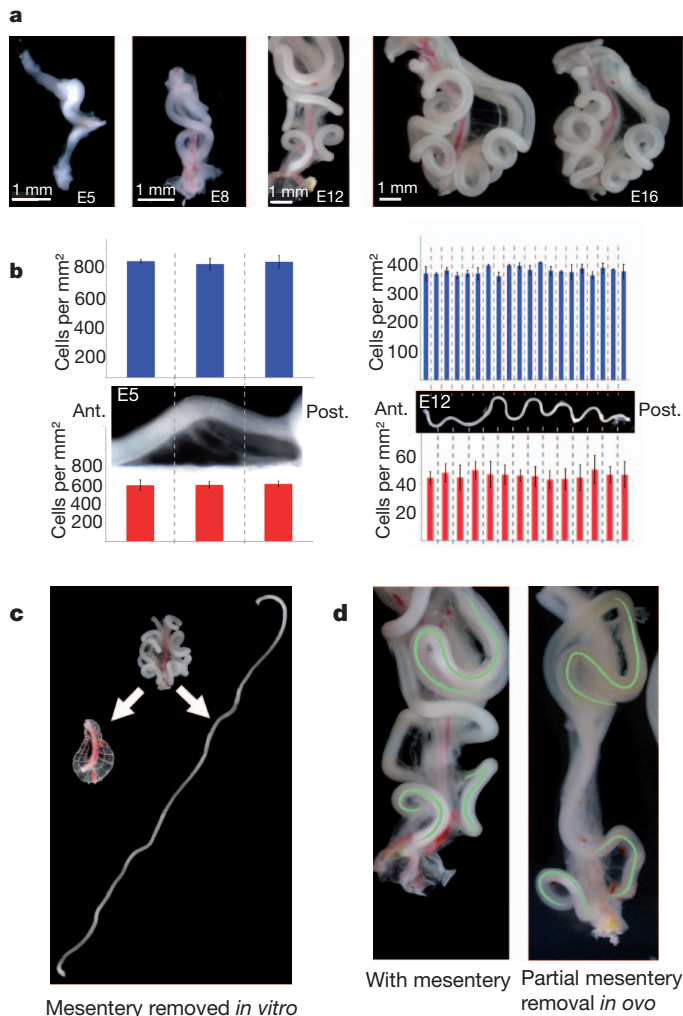
<sup>1</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>3</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>4</sup>Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

<sup>5</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>6</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>7</sup>Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, Massachusetts 02138, USA. †Present addresses: Department of Materials, Polymer Physics, ETH Zürich, 8093 Zürich, Switzerland (T.S.); Department of Molecular Medicine, Cornell University, Ithaca, New York 14853, USA (N.A.K.); Department of Modern Mechanics, USTC-Hefei, Anhui 230027, China (H.L.).

\*These authors contributed equally to this work.





**Figure 1 | Morphology of loops in the chick gut.** **a**, Chick gut at embryonic day 5 (E5), E8, E12 and E16 shows stereotypical looping pattern. **b**, Proliferation in the E5 (left) and E12 (right) gut tubes (blue) and mesentery (red). Each blue bar represents the average number of phospho-H3-positive cells per unit surface in 40 (E5) or 50 (E12) 10- $\mu$ m sections. Each red bar represents the average number of phospho-H3-positive cells per unit surface over six 10- $\mu$ m sections (E5) or in specific regions demarcated by vasculature along the mesentery (E12). The inset images of the chick guts align the proliferation data with the locations of loops (all measurements were made in three or more chick samples). Ant., anterior; post., posterior. Error bars, s.d. **c**, The gut and mesentery before and after surgical separation at E14 show that the mesentery shrinks while the gut tube straightens out almost completely. **d**, The E12 chick gut under normal development with the mesentery (left) and after *in ovo* surgical separation of the mesentery at E4 (right). The gut and mesentery repair their attachment, leading to some regions of normal looping (green). However, a portion of the gut lacks normal loops as a result of disrupting the gut–mesentery interaction over the time these loops would otherwise have developed.

during the formation of the first loop at E5 (Fig. 1b) and later when there were nine loops (E12) (Fig. 1b). We observed consistently uniform proliferation with no significant differences along the rostrocaudal axis of the gut tube, including at loop formation locations and between loops, as well as no observable azimuthal or radial differences in proliferation rates at different cross-sections (Supplementary Fig. 1), consistent with observations that the embryonic gut tube cross-section remains circular along its length.

Because spatial constraints from the body cavity and the gut tube alone cannot explain the reproducible looping, we instead considered the dorsal mesentery, the webbed tissue that attaches the gut tube to

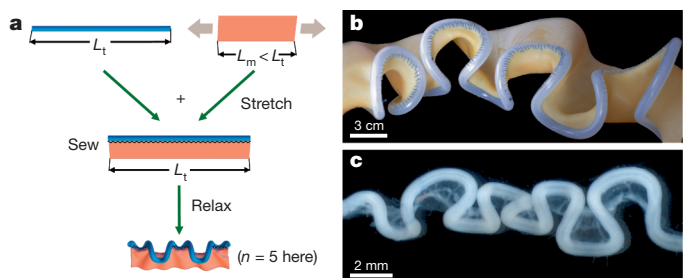
the embryo along its length. As looping morphogenesis is initiated, the dorsal mesentery changes from a thick, asymmetric, multilayer structure to a thin, double-epithelial sheet with no observable left–right asymmetry (Supplementary Fig. 2).

To test whether the dorsal mesentery is integral to the intestinal loops, we separated it from the gut surgically or enzymatically and found that the intestine uncoils into a straight tube, indicating that it was under compression. Simultaneously, the unconstrained dorsal mesentery contracts when freed from the gut tube (Fig. 1c), indicating that this tissue is under tension. Thus the gut–mesentery composite is required to maintain the mature loops in the gut.

To find out whether the dorsal mesentery is also required for the formation of the loops, we surgically separated a portion of the dorsal mesentery from the gut *in ovo*, beginning immediately caudal to the cranial (superior) mesenteric artery (SMA), at day E4, before loops develop. Strikingly, where the mesentery and gut were separated, the intestinal loops failed to form (Fig. 1d) even as normal loops formed in locations rostral and caudal to it (Fig. 1d, green lines). Although we were unable to cut the dorsal SMA *in ovo* during gut loop development, once the loops had matured (E12), surgical dissection of the SMA left the loops intact and in fact highlighted their periodic structure (Fig. 2c). This rules out any possible requirement for the SMA in directing loop structure, and for the vasculature as well, as secondary vessels develop only after the loops themselves have formed.

Although the gut grows uniformly, to investigate whether the mesentery might grow inhomogeneously and thus force the gut to loop at precise locations, we examined the proliferation rate of the mesentery at E5 and at E12. There were no observed differences along the rostrocaudal axis (Fig. 1b), suggesting that the growing mesentery exerts uniform compression along the length of the gut, countered by an equal and opposite tensile reaction on the mesentery from the gut.

Taken together, our observations suggest that uniform differential growth between the gut and the mesentery could be at the origin of loop formation. Because the gut tube is slender, with a length that is much larger than its radius, it responds physically to the differential strain-induced compression from the attached mesentery by bending and looping, while remaining attached to the embryo rostrocaudally. Most importantly, the fact that the gut relaxes to a straight configuration whereas the mesentery relaxes to an almost flat configuration implies that the tissues behave elastically, a fact that will allow us to quantify the process simply.



**Figure 2 | Rubber simulacrum of gut looping morphogenesis.** **a**, To construct the rubber model of looping, a thin rubber sheet (mesentery) was stretched uniformly along its length and then stitched to a straight, unstretched rubber tube (gut) along its boundary; the differential strain mimics the differential growth of the two tissues. The system was then allowed to relax, free of any external forces. **b**, On relaxation, the composite rubber model deformed into a structure very similar to the chick gut (here the thickness of the sheet is 1.3 mm and its Young's modulus is 1.3 MPa, and the radius of the tube is 4.8 mm, its thickness is 2.4 mm and its Young's modulus is 1.1 MPa; see Supplementary Information for details). **c**, Chick gut at E12. The superior mesenteric artery has been cut out (but not the mesentery), allowing the gut to be displayed aligned without altering its loop pattern.

## Physical model of gut looping

To investigate the physical origins of this looping pattern, we developed a simple simulacrum of the gut–mesentery composite using a silicone rubber tube (mimicking the gut) and a thin latex sheet (mimicking the mesentery; see Supplementary Information). The differential strain induced by relative growth between the gut and the mesentery is simulated by extending the latex sheet along its length and stitching it to the wall of the naturally straight, unstretched rubber tube along the edge parallel to the direction of membrane stretching (Fig. 2a). On removing all external loads from the composite system, we observe the spontaneous formation of loops in the tube very similar in shape to the looping patterns seen *in ovo* (Fig. 2b). Varying the differential strain, the thickness of the latex sheet, the radius of the rubber tube and their material properties (Supplementary Information) shows that the wavelength and amplitude of the repeating loops depend only on these measurable parameters.

## Scaling laws for loop period, radius and number

We now quantify the simple physical picture for looping sketched above to derive expressions for the size of a loop, characterized by the contour length,  $\lambda$ , and mean radius of curvature,  $R$ , of a single period (Fig. 3a). The geometry of the growing gut is characterized by the gut's inner and outer radii,  $r_i$  and  $r_o$ , which are much smaller than its increasing length, whereas that of the mesentery is described by its homogeneous thickness,  $h$ , which is much smaller than its other two dimensions. Because the gut tube and mesentery relax to nearly straight, flat states once they are surgically separated, we can model the gut as a one-dimensional elastic filament growing relative to a thin two-dimensional elastic sheet (the mesentery). As the gut length becomes longer than the perimeter of the mesentery to which it is attached, there is a differential strain,  $\varepsilon$ , that compresses the tube axially while extending the periphery of the sheet. When the growth strain is

larger than a critical value,  $\varepsilon_*$ , the straight tube buckles, taking on a wavy shape of characteristic amplitude  $A$  and period  $\lambda \gg A$ . At the onset of buckling, the extensional strain energy of the sheet per wavelength of the pattern is  $U_m \propto E_m \varepsilon_*^2 h \lambda^2$ , where  $E_m$  is the Young's modulus of the mesentery sheet. The bending energy of the tube per wavelength is  $U_t \propto E_t I_t \kappa^2 \lambda$ , where  $\kappa \propto A/\lambda^2$  is the tube curvature,  $I_t \propto r_o^4 - r_i^4$  is the moment of inertia of the tube and  $E_t$  is the Young's modulus of the tube. Using the condition that the in-plane strain in the sheet is  $\varepsilon_* \propto A/\lambda$  and minimizing the sum of the two energies with respect to  $\lambda$  then yields a scaling law for the wavelength of the loop:

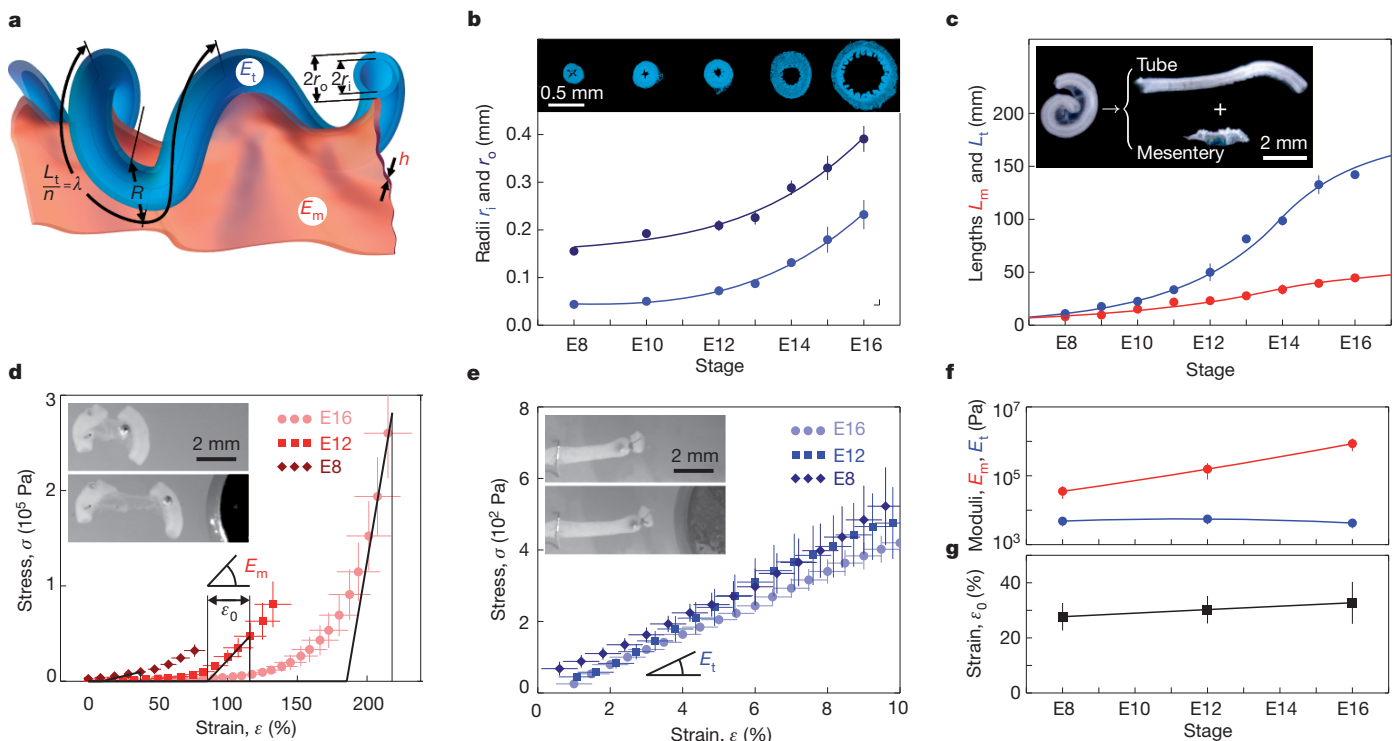
$$\lambda \propto \left( \frac{E_t I_t}{E_m h} \right)^{1/3} \quad (1)$$

The above theory is valid only at the onset of looping and cannot predict the amplitude or radius of a loop. Far from the onset of the instability, at a strain  $\varepsilon = \varepsilon_0 \gg \varepsilon_*$ , we use a torque balance argument to determine the finite radius of the loop. To deform the gut into a loop of radius  $R$ , the elastic torque required is  $T_t \propto E_t I_t / R$  and must balance the torque exerted by the membrane with strain  $\varepsilon_0$  over a width  $w$  and a length  $R$ , that is,  $T_m \propto E_m h w \varepsilon_0 R$ . The width of this strip is the radial distance from the tube over which the peripheral membrane stretching strain is relaxed, and is determined by the relation  $\varepsilon_0 \propto w/(R - w)$ . Balancing the torques, by equating  $T_t$  with  $T_m$ , and assuming that  $\varepsilon_0 < 1$ , yields the scaling law

$$R \propto \left( \frac{E_t I_t}{E_m h \varepsilon_0^2} \right)^{1/3} \quad (2)$$

## Quantitative geometry and biomechanics of chick gut looping

A comparison of the results of our predictions with quantitative experiments requires the measurement of the geometry of the tissues,



**Figure 3 | Geometric and mechanical measurements of chick gut.**

**a**, Parameters involved in the physical model. **b**, Inner ( $r_i$ , light blue) and outer ( $r_o$ , dark blue) tube diameters. Measurements are extracted from DAPI-stained tube cross-section shown in inset. **c**, Tube (length  $L_t$ , blue) and mesentery (length  $L_m$ , red) differential growth. Inset, length measurement of one isolated loop. **d**, Stress versus strain for the mesentery at E8, E12 and E16. For

physiological strains, we use the linearization shown by the black lines, to extract the effective Young's modulus,  $E_m$ , and the effective strain,  $\varepsilon_0$ . **e**, Stress versus strain for the gut tube at E8, E12 and E16. **f**, Mesentery and tube Young's moduli,  $E_m$  (red) and  $E_t$  (blue), at E8, E12 and E16. **g**, Effective differential growth strain,  $\varepsilon_0$ , at E8, E12 and E16. Error bars, s.d.



their elastic properties and the relative strain mismatch at different stages of chick gut development; we chose three stages: E8, E12 and E16 (Fig. 3). The mesentery has a time-varying thickness,  $h$ , which is evaluated from histological cross-section (Supplementary Fig. 2). The inner and outer radii of the gut tube were extracted from 4',6-diamidino-2-phenylindole (DAPI)-stained tube cross-sections (Fig. 3b). The length of the gut tube,  $L_t$ , was measured on the dissected gut. The natural rest length of the periphery of mesentery,  $L_m$ , was measured by cutting out thin strips along the junction with the gut and aligning them unstretched with a ruler (Fig. 3c). The bending stiffness of the gut tube and the stretching stiffness of the mesentery were measured using *in vitro*, uniaxial, low-rate tensile tests, where the load was generated by a magnet applying a calibrated force on a millimetre-size steel ball, attached to one end of a tissue sample that was pinned at the other end. The extension of the sample under load was tracked using video-microscopy to extract its stress ( $\sigma$ )/strain ( $\epsilon$ ) response curve (see Fig. 3d, e, insets, Methods and Supplementary Information).

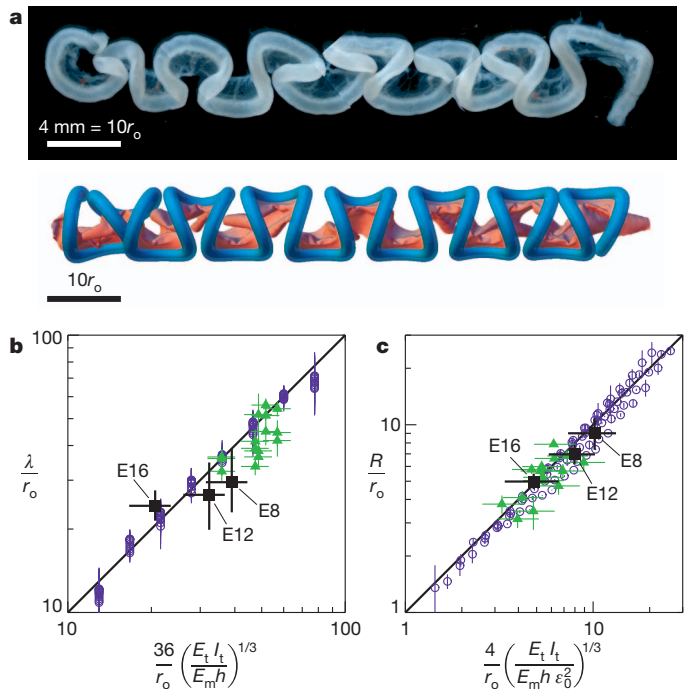
For the mesentery, we observed a nonlinear response curve with a sharp break at a strain  $\epsilon \leq \epsilon_p$ , where  $\epsilon_p = L_t/L_m - 1$  is the physiological strain mismatch, typical of the strain-stiffening seen in biological soft tissues<sup>16</sup>. We define an effective modulus,  $E_m = (d\sigma/d\epsilon)_{\epsilon=\epsilon_p}$ , and strain,  $\epsilon_0 = (\sigma^{-1} d\sigma/d\epsilon)_{\epsilon=\epsilon_p}^{-1}$ , by locally linearizing the response (Fig. 3d) and noting that the membrane has negligible stiffness when  $0 < \epsilon < \epsilon_p - \epsilon_0$ . For the gut, we measured the modulus,  $E_t = \sigma/\epsilon$ , from the linear, low-strain response curve ( $\epsilon < 10\%$ ; Fig. 3e). In Fig. 3f, g, we summarize the variation of  $E_m$ ,  $E_t$  and  $\epsilon_0$  as functions of developmental time. Measurements of the mesentery stiffness at various locations and in various directions did not show significant differences (Supplementary Fig. 4). This confirms the validity of modelling the mesentery and the gut as isotropic, homogeneous material.

The measured biophysical parameters allowed us to create a detailed numerical simulation of gut looping. Because the gut and mesentery grow slowly, inertial effects are unimportant and the composite system is always in mechanical equilibrium. This equilibrium configuration was calculated as follows. The mesentery was modelled as a discrete elastic membrane consisting of a hexagonal lattice of springs with a discrete energy associated with in-plane stretching/shearing deformations as well as out-of-plane bending deformations<sup>17</sup>, relative to the rest length of the springs. The gut was modelled as an equivalent membrane strip (two elements wide) with a discretized energy associated with bending and stretching deformations, and elastic stiffnesses different from those of the membrane. The geometry, mechanical properties and relative growth of the tissues parameterized by  $h$ ,  $L_t$ ,  $E_m$ ,  $E_t$  and  $\epsilon_0$  were all experimentally measured at different time points during development. Given these input parameters, energy minimization for different relative growth strains,  $\epsilon_0$ , yielded predictions for the looping morphology of the gut (Methods and Supplementary Information).

In Fig. 4a, we compare the results of our observation at E16 with numerical simulations. In Fig. 4b, c, we compare our quantitative measurements of the wavelength and radius of curvature of the chick gut at the different measured stages of development (see also Supplementary Fig. 8) with those of both the rubber simulacrum and numerical simulations, as functions of the geometry and elastic moduli of the tube and sheet. Over the strain ranges  $\epsilon_0 \in [0, 1]$  in the simulation (Supplementary Movie 1) and  $\epsilon_0 \in [0.5, 1]$  for the various rubber models, we plot the wavelength,  $\lambda$ , and radius,  $R$ , of the loop and find that they follow the relations

$$\lambda \approx 36 \left( \frac{E_t L_t}{E_m h} \right)^{1/3} \quad (3)$$

$$R \approx 4 \left( \frac{E_t L_t}{E_m h \epsilon_0^2} \right)^{1/3} \quad (4)$$



**Figure 4 | Predictions for loop shape, size and number at three stages in chick gut development.** **a**, Comparisons of the chick gut at E16 (top) with its simulated counterpart (bottom). **b**, Scaled loop contour length,  $\lambda/r_0$ , plotted versus the equivalently scaled expression from equation (3) for the chick gut (black squares), the rubber model (green triangles) and numerical simulations (blue circles). The results are consistent with the scaling law in equation (1). **c**, Scaled loop radius,  $R/r_0$ , plotted versus the equivalently scaled expression from equation (4) for the chick gut, the rubber model, and numerical simulations (symbols are as in **b**). The results are consistent with the scaling law in equation (2). Error bars, s.d.

in accord with our simple scaling laws (equations (1) and (2)). In Table 1, we compare the values of these parameters for the chick gut with the expressions given in equations (3) and (4), and confirm that our model captures the salient properties of the looping patterns with no adjustable parameters, strongly suggesting that the main features of the chick gut looping pattern are established by the simple balance of forces induced by the relative growth between the gut and the mesentery.

### Comparative study of gut looping across species

To test our theory in cases other than the development of the chick gut, we took advantage of the distinct gut looping patterns observed in different avian taxa, which have served as criteria for phylogenetic classification and are thought of as having adaptive significance, independent of bird size.

**Table 1 | Morphometry of chick gut looping pattern**

Stage		$n$	$\lambda$ (mm)	$R$ (mm)
E8	Experimental observation	$2.4 \pm 0.4$	$4.6 \pm 1.0$	$1.4 \pm 0.2$
	Computational model*	$1.8 \pm 0.3$	$6.1 \pm 1.5$	$1.6 \pm 0.3$
E12	Experimental observation	$9.0 \pm 0.5$	$5.6 \pm 1.2$	$1.5 \pm 0.1$
	Computational model†	$7.3 \pm 1.6$	$6.8 \pm 1.6$	$1.7 \pm 0.3$
E16	Experimental observation	$15.0 \pm 0.5$	$9.5 \pm 0.5$	$1.9 \pm 0.1$
	Computational model‡	$17.5 \pm 2.4$	$8.1 \pm 1.9$	$1.9 \pm 0.5$

The observed number of loops ( $n$ ), loop wavelength ( $\lambda$ ) and radius ( $R$ ) for the chick at different stages of gut development, for given geometrical and physical parameters associated with the gut and the mesentery, show that the model predictions are quantitatively consistent with observations.

\*  $L_t = 11.0 \pm 0.5$  mm,  $h = 13.0 \pm 1.5$   $\mu$ m,  $r_0 = 155 \pm 8$   $\mu$ m,  $r_t = 44 \pm 5$   $\mu$ m,  $E_m = 35 \pm 14$  kPa,  $E_t = 4.8 \pm 1.4$  kPa,  $\epsilon_p = 38 \pm 7\%$  and  $\epsilon_0 = 28 \pm 5\%$ .

†  $L_t = 50.0 \pm 8.3$  mm,  $h = 8.0 \pm 1.5$   $\mu$ m,  $r_0 = 209 \pm 12$   $\mu$ m,  $r_t = 72 \pm 9$   $\mu$ m,  $E_m = 156 \pm 78$  kPa,  $E_t = 5.6 \pm 1.7$  kPa,  $\epsilon_p = 116 \pm 19\%$  and  $\epsilon_0 = 30 \pm 5\%$ .

‡  $L_t = 142.1 \pm 3.3$  mm,  $h = 7.1 \pm 1.4$   $\mu$ m,  $r_0 = 391 \pm 27$   $\mu$ m,  $r_t = 232 \pm 31$   $\mu$ m,  $E_m = 861 \pm 344$  kPa,  $E_t = 4.2 \pm 1.3$  kPa,  $\epsilon_p = 218 \pm 15\%$  and  $\epsilon_0 = 33 \pm 8\%$ .

We compared the gut looping patterns of the chick with those of the closely related (but differently sized) quail and those of a songbird, the zebra finch. In Fig. 5a, we see that, as previously described<sup>18,19</sup>, the guts of the chick and the quail are organized almost identically but on different scales, and that the digestive tracts of songbirds and chickens are markedly different. To make the comparison quantitative, we repeated the morphometric and mechanical measurements (Supplementary Information) and used them to generate predictions from our scaling theory and computational model. In all cases, the predicted values of  $\lambda$ ,  $R$  and  $n$  are again in excellent agreement with those observed in embryonic guts of the appropriate species (Fig. 5b, c and Table 2). For instance, we find that although growth strains,  $\varepsilon_p$ , are similar between the chick and the quail, the quail mesentery has a tension,  $E_m h \varepsilon_0$ , approximately five times greater than that in the chick mesentery. Qualitatively, this greater elastic force produces a smaller loop, hence inducing more loops per length and, thus, the same number of loops in the smaller bird. By contrast, most of the geometrical and physical parameters characterizing the developing gut and mesentery in the chick and the zebra finch are different and lead to different looping parameters.

Finally, to test our theoretical model with a non-avian example, we performed a similar set of measurements throughout the course of gut development in mouse embryos. In agreement with our findings from birds, the geometrical and biophysical properties of the developing gut and dorsal mesentery suffice to predict accurately the stereotypical patterns of the mature intestinal loops in mouse embryos (Fig. 5 and

**Table 2 | Morphometry of quail, finch and mouse gut looping patterns**

Species and stage		$n$	$\lambda$ (mm)	$R$ (mm)
Quail E12	Experimental observation	$9.0 \pm 0.7$	$4.6 \pm 0.4$	$1.2 \pm 0.1$
	Computational model*	$10.0 \pm 1.3$	$4.1 \pm 1.0$	$1.2 \pm 0.3$
Finch E13	Experimental observation	$5.5 \pm 0.5$	$3.6 \pm 0.5$	$0.6 \pm 0.3$
	Computational model†	$5.3 \pm 0.8$	$3.7 \pm 0.9$	$0.9 \pm 0.2$
Mouse E16.5	Experimental observation	$6.0 \pm 0.5$	$6.0 \pm 0.7$	$0.7 \pm 0.1$
	Computational model‡	$5.6 \pm 0.8$	$6.4 \pm 1.5$	$1.0 \pm 0.1$

The observed number of loops, loop wavelength and radius for the quail, finch and mouse, for given geometrical and physical parameters associated with the gut and the mesentery, show that the model predictions are quantitatively consistent with observations.

\*  $L_t = 41.3 \pm 0.4$  mm,  $h = 14.9 \pm 1.6$   $\mu$ m,  $r_0 = 248 \pm 13$   $\mu$ m,  $r_i = 154 \pm 12$   $\mu$ m,  $E_m = 515 \pm 206$  kPa,

$E_t = 4.4 \pm 1.3$  kPa,  $\varepsilon_p = 110 \pm 13\%$  and  $\varepsilon_0 = 23 \pm 5\%$ .

†  $L_t = 19.7 \pm 0.8$  mm,  $h = 6.0 \pm 0.6$   $\mu$ m,  $r_0 = 227 \pm 14$   $\mu$ m,  $r_i = 120 \pm 13$   $\mu$ m,  $E_m = 802 \pm 321$  kPa,

$E_t = 2.6 \pm 0.8$  kPa,  $\varepsilon_p = 110 \pm 11\%$  and  $\varepsilon_0 = 32 \pm 5\%$ .

‡  $L_t = 35.9 \pm 0.9$  mm,  $h = 12.3 \pm 1.6$   $\mu$ m,  $r_0 = 270 \pm 16$   $\mu$ m,  $r_i = 178 \pm 14$   $\mu$ m,  $E_m = 94 \pm 37$  kPa,

$E_t = 1.9 \pm 0.9$  kPa,  $\varepsilon_p = 200 \pm 13\%$  and  $\varepsilon_0 = 64 \pm 5\%$ .

Table 2). The mouse gut is notably characterized by softer tissues and higher mismatch strain, producing tightly coiled loops, as seen in Fig. 5a. The physiological stresses in the mesentery fall in the same range (Supplementary Information) in all the species investigated in this study, suggesting that both growth and the properties of tissues might be regulated by mechanical feedback.

## Discussion

The developing intestine is a simple, elongated, tubular structure that is stereotypically and reproducibly folded into a compact organ through the process of looping morphogenesis. Using a combination of quantitative experiments, computations and scaling arguments, we have shown that the associated looping patterns are quantitatively determined by the differential growth between the gut tube and the dorsal mesentery and by their geometric and elastic properties, both within individual organisms and across species. We thus bring a quantitative biomechanical perspective to the mostly metaphoric arguments in *On Growth and Form*<sup>2</sup>.

The simplicity of the mechanical origin in the diversity in gut looping patterns, long associated with the adaptive significance of the distinct diets and gut residence times of different animals<sup>18</sup>, also suggests that because it is sufficient to modulate the uniform tissue growth rates, tissue geometry and elasticity of the gut–mesentery system to change these patterns, this is the minimal set of properties on which selection has acted to achieve the looping patterns found in nature.

Identification of the relevant cellular parameters influencing gut morphogenesis opens the door to future studies of the genes involved in controlling cell proliferation and matrix formation in space and time, and sets the stage to understanding the processes by which biochemical and biophysical events across scales conspire to drive the developmental regulation of growing tissues.

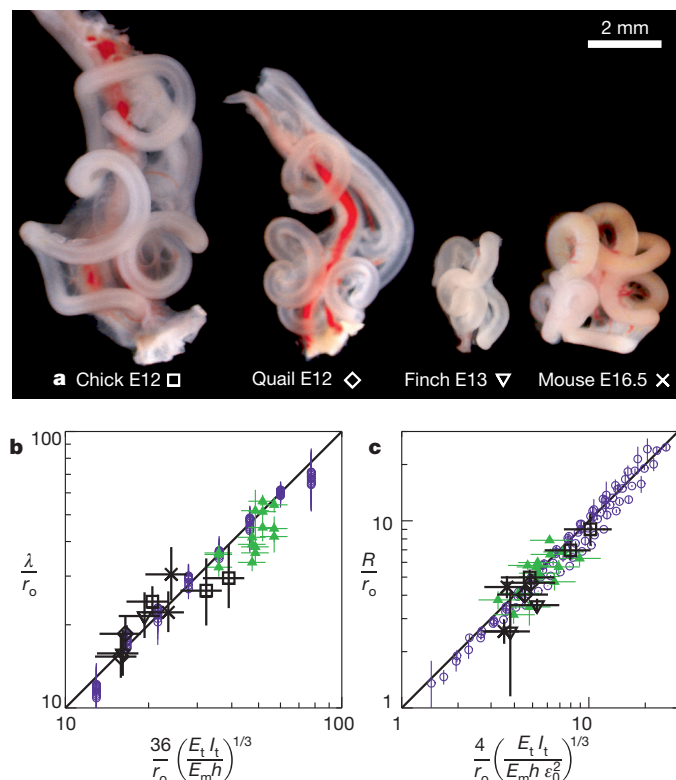
## METHODS SUMMARY

**Embryos.** Fertile chick eggs (White Leghorn eggs) were obtained from commercial sources. Fertile zebra finch eggs were provided by the laboratory of T. Gardner at Boston University. Fertile Japanese quail eggs were obtained from Strickland Game Bird. All eggs were incubated at 37.5 °C and staged following ref. 20. Mouse embryos were collected from staged pregnant females (Charles River Laboratories).

**Immunohistochemistry and histology.** Small intestines were collected from chick embryos at desired stages and fixed in 4% paraformaldehyde in PBS and embedded in paraffin wax. Immunohistochemistry and histology was performed on 10- $\mu$ m transverse sections of the gut tube.

**In ovo gut surgeries.** The gut tube and the dorsal mesentery were separated *in ovo* at stage 23–25 by using a pulled glass needle to cut the connection between the two tissues. Embryos were re-incubated until E12, when they were collected to examine the resulting looping pattern.

**Mechanical properties of gut and mesentery tissue.** The force,  $F(d)$ , between a permanent magnet (The Magnet Source) and millimetre-size steel balls (New England Miniature Ball Corp.), separated by a distance  $d$ , was calculated from the damped motion of the ball rising in glycerol with the magnet lowered from



**Figure 5 | Comparative predictions for looping parameters across species.** **a**, Gut looping patterns in the chick, quail, finch and mouse (to scale) show qualitative similarities in the shape of the loops, although the size and number of loops vary substantially. **b**, Comparison of the scaled loop contour length,  $\lambda/r_0$ , with the equivalently scaled expression from equation (3) shows that our results are consistent with the scaling law in equation (1) across species. Black symbols are for the animals shown in **a**, other symbols are the same as in Fig. 4b. **c**, Comparison of the scaled loop radius,  $R/r_0$ , with the equivalently scaled expression from equation (4) shows that our results are consistent with the scaling law in equation (2) across species (symbols are as in **b**). In **b** and **c**, points are reported for chick at E8, E12 and E16; quail at E12 and E15; finch at E10 and E13; and mouse at E14.5 and E16.5. Error bars, s.d.



above. For  $2\text{ mm} < d < 8\text{ mm}$ , the range used in the subsequent measurements,  $F(d)$  ranges from  $1\text{ }\mu\text{N}$  to  $1\text{ mN}$ . We then surgically extracted strips of the mesentery and sections of the gut tube from fresh animal embryos. A steel bead was attached at one end of the sample, by either gluing it onto the strip or sealing it into the tube. With the other end of the sample pinned to an agarose gel, the magnet was moved closer to stretch the sample while  $d$  and the extension,  $L$ , of the sample were tracked by video. We then calculated the stress,  $\sigma = F(d)/A_0$ , and the strain,  $\varepsilon = L/L_0 - 1$ , where  $L_0$  and  $A_0$  are respectively the length and the cross-sectional area of the sample at rest. All dissections, manipulations and tensile tests occurred in Ringer buffer (Sigma Aldrich), and within hours after the surgery.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 23 January; accepted 7 June 2011.**

- His, W. *Anatomie Menschlicher Embryonen* (Vogel, 1880).
- Thompson, D. W. *On Growth and Form* (Cambridge Univ. Press, 1917).
- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Metzger, R. J. & Krasnow, M. A. Genetic control of branching morphogenesis. *Science* **284**, 1635–1639 (1999).
- Hufnagel, L. *et al.* On the mechanism of wing size determination in fly development. *Proc. Natl Acad. Sci. USA* **104**, 3835–3840 (2007).
- Beloussov, L. V. *et al.* Mechanical stresses in embryonic tissues: patterns, morphogenetic role, and involvement in regulatory feedback. *Int. Rev. Cytol.* **150**, 1–34 (1994).
- Taber, L. A. Biomechanics of cardiovascular development. *Annu. Rev. Biomed. Eng.* **3**, 1–25 (2001).
- Salazar-Ciudad, I. & Jernvall, J. A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**, 583–586 (2010).
- Hamant, O. *et al.* Developmental patterning by mechanical signals in *Arabidopsis*. *Science* **322**, 1650–1655 (2008).
- Forgacs, G. & Newman, S. *Biological Physics of the Developing Embryo* (Cambridge Univ. Press, 2005).
- Schoenwolf, G. *et al.* *Larsen's Human Embryology* Ch. 14 (Elsevier Health Sciences, 2008).
- Kurpios, N. A. *et al.* The direction of gut looping is established by changes in the extracellular matrix and in cell:cell adhesion. *Proc. Natl Acad. Sci. USA* **105**, 8499–8506 (2008).
- Davis, N. M. *et al.* The chirality of gut rotation derives from left-right asymmetric changes in the architecture of the dorsal mesentery. *Dev. Cell* **15**, 134–145 (2008).
- Hecksher-Sorensen, J. *et al.* The splanchnic mesodermal plate directs spleen and pancreatic laterality, and is regulated by *Bapx1/Nkx3.2*. *Development* **131**, 4665–4675 (2004).
- Kleinman, R. E. *et al.* *Walker's Pediatric Gastrointestinal Disease* 207–216 (Decker, 2008).
- Fung, Y. C. *Biomechanics: Mechanical Properties of Living Tissues* 2nd edn, 242–320 (Springer, 2004).
- Liang, H. & Mahadevan, L. The shape of a long leaf. *Proc. Natl Acad. Sci. USA* **106**, 22049–22054 (2009).
- Beddard, F. E. *The Structure and Classification of Birds* (Longmans, Green and Co., 1898).
- Mitchell, C. P. On the intestinal tract of birds. *Proc. Zool. Soc. Lond.* **64**, 136–159 (1896).
- Hamburger, H. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Exp. Morphol.* **88**, 49–92 (1951).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Prum for pointing out to us the literature on avian intestines, and the Harvard NSF MRSEC, the MacArthur Foundation (L.M.) and NIH RO1 HD047360 (C.J.T.) for support.

**Author Contributions** C.J.T., N.A.K. and L.M. designed the research with additional contributions from T.S. and A.E.S.; T.S. (biophysical and computational experiments, data analysis), N.A.K. (biological experiments), A.E.S. (biological and biophysical experiments) and L.M. (physical mechanism, physical/mathematical model, scaling theory) did the research; P.F. (stitched physical model) and H.L. (built computational model) contributed tools; and T.S., N.A.K., L.M. and C.J.T. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.M. (lm@seas.harvard.edu).

## METHODS

**Embryos.** Fertile chick eggs (White Leghorn eggs) were obtained from commercial sources. Fertile zebra finch eggs were provided by the laboratory of T. Gardner at Boston University. Fertile Japanese quail eggs were obtained from Strickland Game Bird. All eggs were incubated at 37.5 °C and staged following ref. 20. Mouse embryos were collected from staged pregnant females (Charles River Laboratories).

**Immunohistochemistry and histology.** Small intestines were collected from chick embryos at desired stages and fixed in 4% paraformaldehyde in PBS and embedded in paraffin wax, allowing for 10-µm transverse sections of the gut tube. Fast green staining was performed as described in ref. 20. Immunohistochemistry was performed with rabbit polyclonal antiphospho-H3 (1:100) (Millipore) overnight at 4 °C in PBS containing 3% goat serum and 0.1% Triton X-100. Sections were next incubated with Alexa Fluor 594 goat anti-rabbit secondary antibody (1:300; Molecular Probes) for 1 h at room temperature (25 °C). DAPI (Molecular Probes) was used as a nuclear counterstain and to determine the inner and outer radii of the gut tube.

**In ovo gut surgeries.** The gut tube and the dorsal mesentery were separated *in ovo* at stage 23–25 (ref. 20) by using a pulled glass needle to cut the connection between the two tissues. Most, but not all, of the connection was ablated as care was taken to avoid puncturing the dorsal aorta, which runs over the gut tube and dorsal mesentery at this stage. Embryos were re-incubated until E12, when they were collected to examine the resulting looping pattern.

**Physical simulacrum using rubber.** We used wide strips of elastic rubber sheeting (McMaster-Carr) of various thicknesses. Each strip was held stretched in one direction at the desired extension using clamps, and a silicon rubber tube (NewAge Industries) was stitched to the sheet using sewing thread (Supplementary Information).

**Calibration of the magnetic force.** The attractive interaction between a permanent disc magnet (commercial grade, axially magnetized, neodymium Nd-Fe-B; The Magnet Source) and high-precision steel balls (AISI 440C stainless steel, radii  $r_b = 0.122, 0.253$  and  $0.398$  mm; New England Miniature Ball Corp.) was calibrated using a ‘falling-ball viscometer’ geometry: immersed in a tube filled with pure glycerol, the magnet is brought closer to the ball from above, and the ball consequently rises (we ensured that all materials used to manipulate the beads and the magnet during the measurements, calibration and tensile tests had no magnetic susceptibility). The force exerted by the magnet is balanced by gravity, drag and inertia. At low Reynolds number, drag force and inertia can be measured from the ball trajectory that is extracted using video tracking (see details in the Supplementary Information). We can then calculate the attractive force,  $F(d)$ , between the magnet and the ball as a function of their separation distance,  $d$ . We report our results in Supplementary Fig. 3 and Supplementary Table 1. Notably, for the distances,  $2 \text{ mm} < d < 8 \text{ mm}$ , used in the tissue tensile test, the force ranges from  $1 \text{ µN}$  to  $1 \text{ mN}$ .

**Measurements of tissue mechanical properties.** We surgically dissected fragments of the mesentery and of the gut tube from live embryos. Samples of the mesentery were cut out to leave a well-defined, millimetre-width strip with principal axis either perpendicular to the tube (radial measurement) or parallel to the tube (tangential measurement). For mesentery fragments, the steel beads were glued using synthetic glue (Instant Krazy Glue) at one end of the tissue strip. The other end was pinned to an agarose gel layer. During the dissection of the sample, we kept sections of the tube or of the superior mesenteric artery to provide convenient handles to which to attach the bead (see Fig. 3d and Supplementary

Fig. 4, where sections of the tube are visible). For gut tube fragments, the steel beads were inserted into the tube and secured by tying the lumen using a hair with an overhand knot. The other end of the tube was held on the agarose gel using a horseshoe pin (Fig. 3e and Supplementary Fig. 5). All dissections, manipulations and tensile tests were performed in Ringer buffer (Sigma Aldrich), and the measurements were made within a few hours of the dissection.

The magnet was attached to a plastic arm held on a micrometric translation stage, and moved closer to the sample on the agarose gel (Supplementary Fig. 4a). The magnet attracted the steel bead and stretched the sample in a controlled fashion (Supplementary Movie 2). The tensile tests were video-recorded to track the extension,  $L$ , of the sample and the bead–magnet distance,  $d$ , and were run as follows. The sample was first pre-conditioned by stretching it once to an extension ratio greater than one, after which the magnet was removed to let the sample relax to its rest length,  $L_0$  (Supplementary Movie 2), at which stage we measured the rest width,  $w_0$ , of the mesentery sample. The magnet was then moved back towards the sample in a stepwise manner. At each step, the sample stretched and was confirmed visually to have reached equilibrium extension before the next step was taken. We thus effectively measured the static elasticity of the tissue, in terms of the nominal stress,  $F(d)/A_0$ , produced by a nominal strain,  $\varepsilon = L/L_0 - 1$ . Here  $A_0$  is the cross-section of the sample at rest:  $A_0 = w_0(1 + \varepsilon_p)^{1/2}h$  for the mesentery, by virtue of material incompressibility, and  $A_0 = \pi(r_o^2 - r_i^2)$  for the tube.

We verified that this method gives reproducible results, and we found that the same-sample variations were less than sample-to-sample variations that we measured at about 50%. Several stress–strain response curves, corresponding to samples of mesentery and gut tubes extracted from different chick E16 embryos are shown in Supplementary Figs 4c and 5b and indicate the level of reproducibility.

**Computational model.** The mesentery was modelled as a hexagonal lattice of linear springs with rest length  $a_m$ , whose discrete energy

$$F_m = \frac{\sqrt{3}E_m h}{4} \sum_{ij} (r_{ij} - a_m)^2 + \frac{E_m h^3}{12\sqrt{3}} \sum_{\alpha\beta} (\mathbf{n}_\alpha - \mathbf{n}_\beta)^2$$

accounts for in-plane stretching (first term, where  $r_{ij}$  is the spring length between nodes  $i$  and  $j$ ) and out-of-plane bending (second term, where  $\mathbf{n}_\alpha$  is the unit normal vector to the triangular facet), and tends to the energy of an elastic membrane of thickness  $h$  and Young’s modulus  $E_m$  as  $a_m \rightarrow 0$  (ref. 17). The gut tube was modelled with a similar, but two-element-wide, lattice of springs with rest length  $a_t$ . The discrete energy

$$F_t = \frac{5.6\sqrt{3}E_t I_t}{8r_o^3} \sum_{ij} (r_{ij} - a_t)^2 + \frac{E_t I_t}{2r_o\sqrt{3}} \sum_{\alpha\beta} (\mathbf{n}_\alpha - \mathbf{n}_\beta)^2$$

of this strip also contains in- and out-of-plane deformations terms (first and second terms, respectively), which are chosen such that the in- and out-of-plane bending stiffnesses both converge to  $E_t I_t$  for a tube of outer radius  $r_o$  (see details in Supplementary Information). At various time points in the development of the gut (E8, E12 and E16), the parameters  $h$ ,  $I_t$ ,  $E_m$ ,  $E_t$  and  $\varepsilon_0$  are all experimentally measured and input into the energy, with the relative growth,  $\varepsilon_0 = a_m/a_t - 1$ , imposing the mismatch strain between the membrane and the tube attached to it. Then the energy  $F_m + F_t$  is minimized using a damped molecular dynamics algorithm<sup>17</sup>, to yield the equilibrium configuration of the gut–mesentery composite system.



# A critical role for TCF-1 in T-lineage specification and differentiation

Brittany Nicole Weber<sup>1\*</sup>, Anthony Wei-Shine Chi<sup>1\*</sup>, Alejandro Chavez<sup>1</sup>, Yumi Yashiro-Ohtani<sup>1</sup>, Qi Yang<sup>1</sup>, Olga Shestova<sup>1</sup> & Avinash Bhandoola<sup>1</sup>

**The vertebrate thymus provides an inductive environment for T-cell development. Within the mouse thymus, Notch signals are indispensable for imposing the T-cell fate on multipotential haematopoietic progenitors, but the downstream effectors that impart T-lineage specification and commitment are not well understood. Here we show that a transcription factor, T-cell factor 1 (TCF-1; also known as transcription factor 7, T-cell specific, TCF7), is a critical regulator in T-cell specification. TCF-1 is highly expressed in the earliest thymic progenitors, and its expression is upregulated by Notch signals. Most importantly, when TCF-1 is forcibly expressed in bone marrow (BM) progenitors, it drives the development of T-lineage cells in the absence of T-inductive Notch1 signals. Further characterization of these TCF-1-induced cells revealed expression of many T-lineage genes, including T-cell-specific transcription factors *Gata3* and *Bcl11b*, and components of the T-cell receptor. Our data suggest a model where Notch signals induce TCF-1, and TCF-1 in turn imprints the T-cell fate by upregulating expression of T-cell essential genes.**

Within the thymus, Notch1 signals drive development through sequential steps during which alternative lineage potentials are lost and T-lineage-specific gene expression (specification) occurs<sup>1–4</sup>. Notch is necessary for early T-cell development but its downstream effectors remain unclear<sup>5–7</sup>. We found that the high mobility group (HMG) box transcription factor, TCF-1, is highly upregulated in early thymic progenitors (ETPs; Fig. 1a). Indeed, TCF-1 expression is upregulated when progenitors are exposed to Notch1 signals<sup>8</sup>.

## TCF-1 in normal T-lymphopoiesis

TCF-1 deficiency greatly reduces thymic cellularity but does not abrogate T-cell development<sup>9–11</sup> (Supplementary Fig. 1). When *TCF-1*<sup>−/−</sup> progenitors were assessed in the absence of competition in irradiated mice, small numbers of T-lineage cells developed (Supplementary Fig. 2a). The related transcription factor LEF-1 can compensate for TCF-1 (ref. 12); consistently, *TCF-1*<sup>−/−</sup> DN3 cells showed elevated LEF-1 expression (Supplementary Fig. 2d). To examine more rigorously the requirement for TCF-1 in early progenitors, we placed *TCF-1*<sup>−/−</sup> progenitors in competition with wild-type cells in mixed BM chimaeras. *TCF-1*<sup>−/−</sup> progenitors reconstituted BM progenitor populations but were defective in generating ETPs, and downstream thymic populations were almost entirely absent (Fig. 1b, c). These data indicated a marked requirement for TCF-1 at very early stages of T-cell development, which was clearly revealed when TCF-1-deficient progenitors were placed in competition with TCF-1-sufficient cells.

To elucidate more precisely the role of TCF-1 in early T-cell development, we used stromal cells expressing Notch ligands (OP9-DL4 or OP9-DL1). In this system, haematopoietic progenitors that respond to Notch signals differentiate into immature Thy1<sup>+</sup>CD25<sup>+</sup> T-lineage cells<sup>13,14</sup>. Both *TCF-1*<sup>+/−</sup> and *TCF-1*<sup>−/−</sup> Lin<sup>−</sup>Scal<sup>+</sup>Kit<sup>+</sup> (LSK) progenitors generated myeloid and B-lineage cells on control OP9 stroma and these fates were appropriately inhibited when progenitors were signalled through Notch. On OP9-DL1 stroma, however, *TCF-1*<sup>−/−</sup> progenitors failed to give rise to T-lineage cells (Fig. 1d, e), even when the survival factor Bcl-xL was ectopically

expressed (Supplementary Fig. 3a). Hence TCF-1 is dispensable for initial Notch1-mediated inhibition of alternative fates but is involved in promoting the T-cell fate.

To better examine the requirement for TCF-1 in promoting T-cell development, we cultured *TCF-1*<sup>−/−</sup> and *TCF-1*<sup>+/+</sup> lymphoid-primed multipotent progenitors (LMPPs) on OP9-DL4 for 4 days and performed global gene expression analysis on *TCF-1*<sup>−/−</sup> and *TCF-1*<sup>+/+</sup> lineage-negative precursors as well as *TCF-1*<sup>+/+</sup> Thy1<sup>+</sup>CD25<sup>+</sup> T-lineage cells. We found that *TCF-1*<sup>−/−</sup> progenitors failed to upregulate expression of many T-lineage genes (Fig. 1f, g). Both *TCF-1*<sup>+/+</sup> and *TCF-1*<sup>−/−</sup> progenitors upregulated expression of Notch target genes *Deltex1* (also known as *Dtx1*) and *Hes1* (Supplementary Fig. 4), confirming that TCF-1-deficient progenitors sense Notch signals, but cannot upregulate expression of T-cell genes.

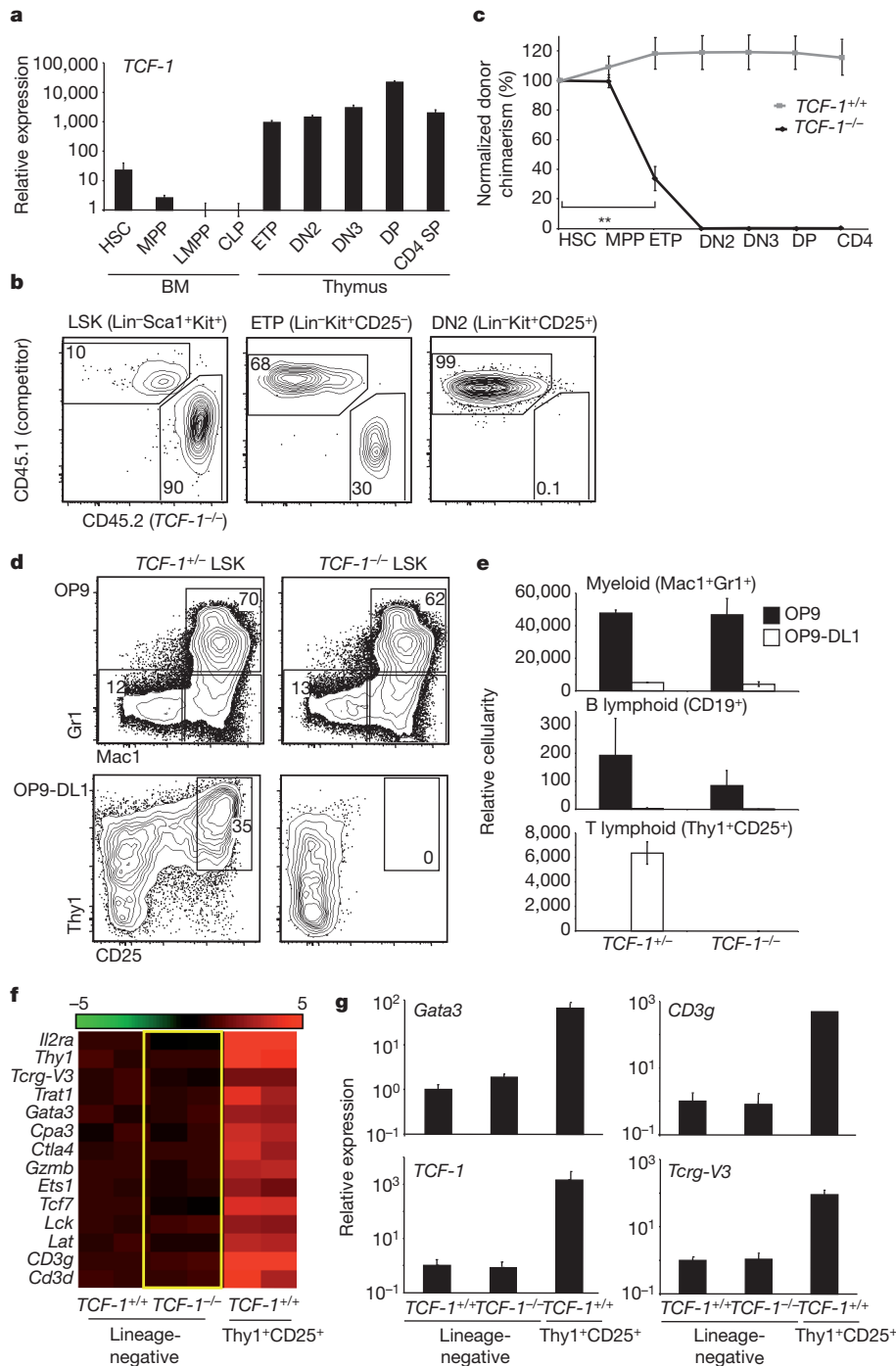
## TCF-1 drives early T-cell development

To investigate the possibility that TCF-1 initiates T-lineage gene expression, we expressed human TCF-1 ectopically in LMPPs. T-lineage cells were observed from TCF-1-expressing wild-type LMPPs on OP9-DL4 stroma, as expected, and ectopic TCF-1 rescued T-cell development from *TCF-1*<sup>−/−</sup> progenitors (Fig. 2a and Supplementary Fig. 3b). Ectopic TCF-1 and Notch1 signals together enhanced T-cell development (Supplementary Fig. 5). However, when TCF-1-expressing progenitors were placed on OP9 stromal cells lacking Notch ligands, we also observed the development of T-lineage cells; this population was absent from progenitors transduced with control virus cultured on OP9 stroma (Fig. 2a). Ectopic expression of TCF-1 also efficiently inhibited the development of the B lineage but not of myeloid cells (Fig. 2b). However, because Notch signals efficiently inhibited the development of B cells from *TCF-1*<sup>−/−</sup> progenitors (Fig. 1e), other mechanisms apart from TCF-1 to enforce lineage commitment must exist.

We further investigated the TCF-1-mediated generation of Thy1<sup>+</sup>CD25<sup>+</sup> cells on OP9 stroma. These cells appeared early and expanded in number over time. They expressed surface markers of

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, 3620 Hamilton Walk Philadelphia, Pennsylvania 19104, USA.

\*These authors contributed equally to this work.



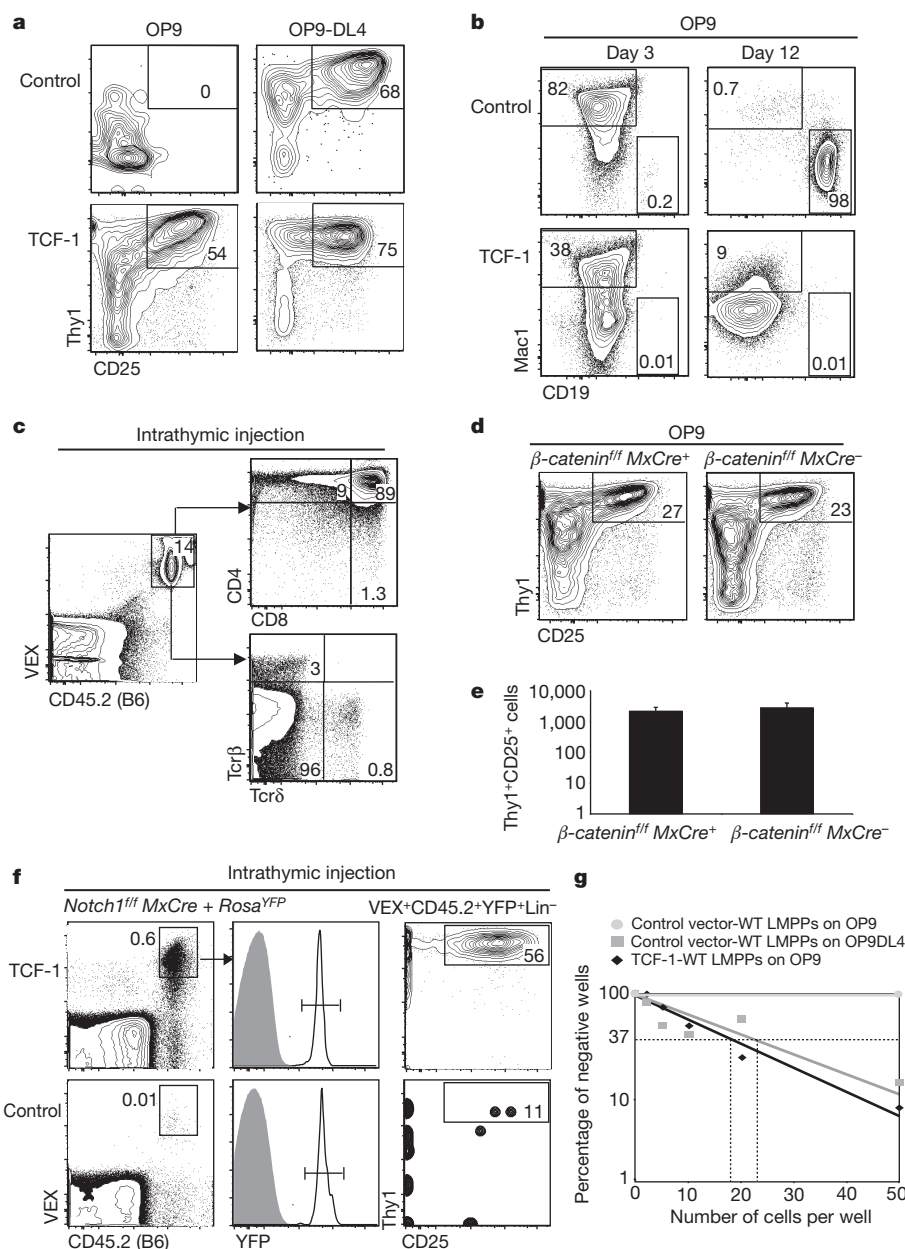
**Figure 1 | TCF-1 is necessary for early T-lineage development and specification.** **a**, *TCF-1* gene expression in BM, thymic progenitors and T-cells. Expression is shown relative to 18S RNA and LMPP. CLP, common lymphoid progenitor. **b**, Mixed BM chimeras were generated using *TCF-1*<sup>-/-</sup> BM and wild-type BM. **c**, Chimerism of *TCF-1*<sup>-/-</sup> cells was normalized to HSC (four mice per group; three independent experiments), \*\**P* < 0.005. **d**, *TCF-1*<sup>+/+</sup> and *TCF-1*<sup>-/-</sup> LSK progenitors were seeded onto OP9 and OP9-DL1 stroma and analysed for myeloid (Mac1<sup>+</sup>Gr1<sup>+</sup>) and T development (Thy1<sup>+</sup>CD25<sup>+</sup>). **e**, Cellularity of day 6 cultures, including B (CD19<sup>+</sup>) is shown. **f**, *TCF-1*<sup>-/-</sup> and

double-negative (DN) 2 and DN3 pro-T cell stages. A different retroviral vector that expresses TCF-1 at lower levels failed to generate Thy1<sup>+</sup>CD25<sup>+</sup> cells, indicating a threshold level of TCF-1 expression is necessary. The generation of Thy1<sup>+</sup>CD25<sup>+</sup> cells was unaffected by inhibitors of Notch signalling (Supplementary Fig. 6). When injected intrathymically, these cells completed T-cell differentiation, reconstituting both αβ and γδ T-cell lineages (Fig. 2c).

*TCF-1*<sup>+/+</sup> LMPPs were seeded onto OP9-DL1 and lineage-negative cells from *TCF-1*<sup>+/+</sup> and *TCF-1*<sup>-/-</sup> cultures were harvested at day 4 for gene expression. The right side of panel corresponds to T lineage cells made from normal progenitors at day 4 in culture. Lineage-negative cells from these early cultures retain progenitor activity<sup>8</sup>. Heat map shows a selection of T-lineage genes with expression increased more than twofold compared with *TCF-1*<sup>+/+</sup> lineage-negative cells and represents the log<sub>2</sub> value of normalized signal level. Rows represent two independent samples for each population. **g**, QRT-PCR validation of selected genes. All error bars are means ± s.e.m.

TCF-1 can function with β-catenin to mediate canonical Wnt signalling; however, deletion of β-catenin does not affect T-cell development<sup>15,16</sup>. Consistently, the generation of Thy1<sup>+</sup>CD25<sup>+</sup> cells was unaffected by deletion of β-catenin (Fig. 2d, e). Furthermore, ectopic expression of a small molecule inhibitor of β and γ-catenin, ICAT<sup>17</sup>, had no effect on the generation of TCF-1-expressing Thy1<sup>+</sup>CD25<sup>+</sup> cells, demonstrating that TCF-1 is not acting as an





**Figure 2 | Ectopic expression of TCF-1 elicits T-lineage cells *in vitro*.**

**a**, Wild-type LMPPs were transduced with control murine stem cell virus containing Violet-Excited GFP (MSCV-VEX) or MSCV containing human TCF-1 (MSCV-TCF-1-VEX). Transduced cells were isolated by cell sorting, and seeded onto OP9 or OP9-DL4. Plots are gated on VEX<sup>+</sup>CD45.2<sup>+</sup> Mac1<sup>-</sup>Gr1<sup>-</sup> cells, shown on day 12. **b**, On OP9 stroma, TCF-1-expressing progenitors gave rise to myeloid cells (Mac1<sup>+</sup>Gr1<sup>+</sup>), shown on day 3, but TCF-1 inhibited the development of CD19<sup>+</sup> B cells, shown on day 12. **c**, TCF-1-expressing Thy1<sup>+</sup>CD25<sup>+</sup> cells were isolated from OP9 cultures after 8 days and injected intrathymically into congenic recipients. Shown is 19 days post injection. **d**, β-catenin<sup>+/f</sup>MxCre<sup>+</sup> and control mice were induced with poly(I:C) and LMPPs were isolated, transduced with MSCV-TCF-1-VEX, and 2,000 transduced cells were seeded per well on OP9 stroma and analysed at day 7, 10

and 12. Plots are gated on VEX<sup>+</sup>CD45.2<sup>+</sup> Mac1<sup>-</sup>Gr1<sup>-</sup> cells, shown on day 10. **e**, Relative cellularity of day 7 cultures. Results represent triplicates ± s.d. **f**, Notch1<sup>+/f</sup>MxCre<sup>+</sup>Rosa<sup>YFP/+</sup> mice were induced with poly(I:C) and YFP<sup>+</sup> LSK progenitors were isolated and transduced with TCF-1 or vector control and injected intrathymically into sublethally irradiated recipients. Shown is day 10 analysis, two independent experiments, 4–6 mice per experiment. Frequency of donor-derived TCF-1-expressing Thy1<sup>+</sup>CD25<sup>+</sup> cells compared to control,  $P = 0.03$ . **g**, Limiting dilution analysis was performed on TCF-1-expressing LMPPs grown on OP9 stroma; this was compared to LMPPs grown on OP9-DL4 stroma. Frequencies of lineage-competent cells were similar. (TCF-1-expressing Thy1<sup>+</sup>CD25<sup>+</sup> lineage on OP9: 1 in 17 (95% confidence interval 1 in 12–26), control Thy1<sup>+</sup>CD25<sup>+</sup> lineage on OP9-DL4: 1 in 23 (95% confidence interval 1 in 15–34).)

effector of canonical Wnt signalling in early T-cell development (Supplementary Fig. 7). Ectopic expression of TCF-1 in long-term haematopoietic stem cells (HSCs), but not in myelo-erythroid progenitors, resulted in development of T-lineage cells on OP9 stroma, indicating that TCF-1-directed T-lineage development is not restricted to lymphoid-biased progenitors (Supplementary Fig. 8a), and requires factors absent from committed myelo-erythroid progenitors (Supplementary Fig. 8b). These results indicate that TCF-1

is sufficient to induce the development of primitive haematopoietic progenitors into cells phenotypically and functionally resembling early T-cell precursors.

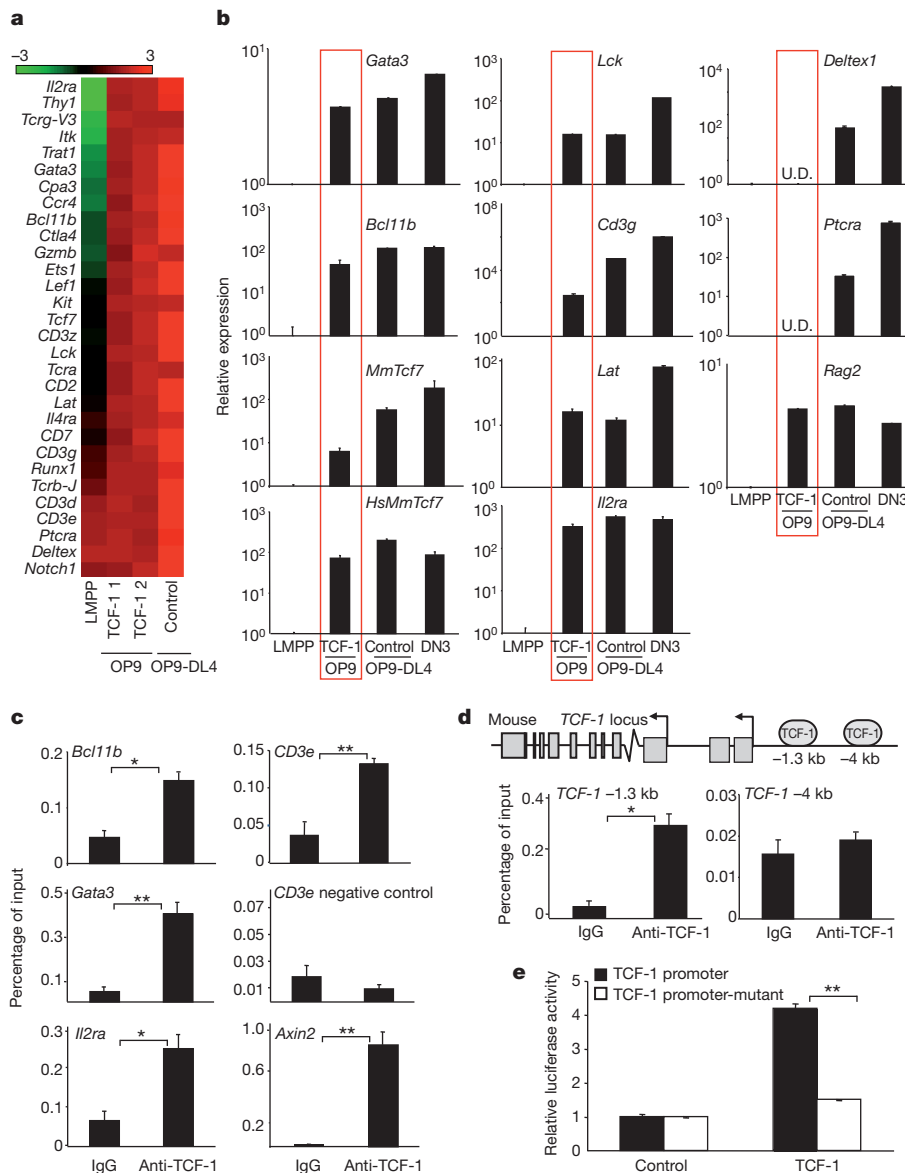
We studied the effects of ectopic expression of TCF-1 *in vivo*. When TCF-1-expressing progenitors were injected intravenously into irradiated mice, we did not observe T-cell leukaemia, unlike forced expression of intracellular Notch1 (ICN1) (Supplementary Fig. 9)<sup>18</sup>. These data signify that key gene targets of ICN1 that control growth and oncogenesis

are not similarly triggered by TCF-1. We next intrathymically injected TCF-1-expressing or control-vector-expressing progenitors from *Notch1<sup>fl/fl</sup>MxCre<sup>+</sup>Rosa<sup>YFP/+</sup>* mice that had been induced with polyinosinic:polycytidylic acid (poly(I:C)). TCF-1-expressing progenitors lacking *Notch1* gave rise to DN2/3-like *Thy1<sup>+</sup>CD25<sup>+</sup>* cells, whereas control progenitors lacking *Notch1* developed into B-lineage cells (Fig. 2d and Supplementary Fig. 10). Therefore forced expression of TCF-1 can drive early T-cell development in the absence of *Notch1* signals in the thymus.

To investigate the frequency of TCF-1-expressing LMPPs able to give rise to T-lineage cells, we performed limiting dilution analysis with TCF-1-expressing LMPPs on OP9 stromal cells and vector-control-expressing LMPPs on OP9-DL4. The frequencies of T-lineage cells developing in these cultures were similar (Fig. 2e). Thus, ectopic TCF-1 generates phenotypic T-cell precursors with frequencies comparable to *Notch*.

### TCF-1 directs T-lineage specification

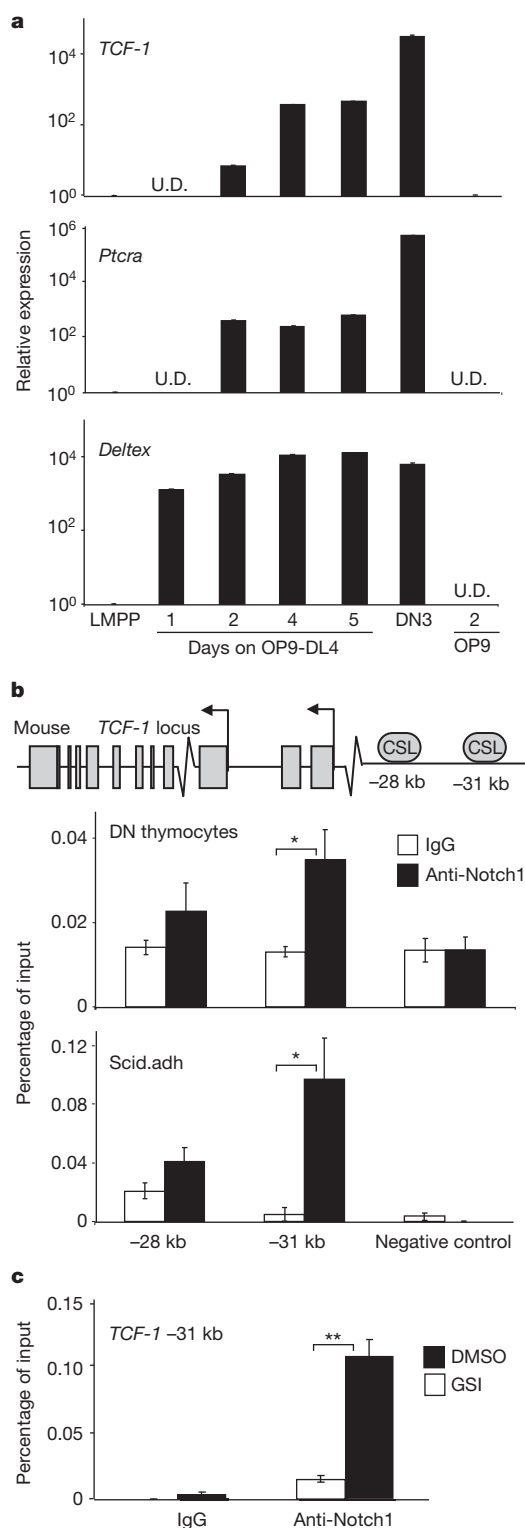
To understand whether TCF-1 is sufficient to direct a program of T-lineage-specific gene expression, we performed global gene expression analysis on TCF-1-expressing *Thy1<sup>+</sup>CD25<sup>+</sup>* T-lineage cells that developed on OP9 stroma. We found upregulated expression of many T-cell genes, including transcription factors *Gata3* and *Bcl11b*, and T-cell structural genes including components of the T cell receptor (Fig. 3a). Established direct *Notch1* gene targets such as *Ptcr*a and *Deltex1* (ref. 19) failed to be upregulated, confirming that these T-lineage cells arose independently of *Notch1* signals (Fig. 3b). Quantitative PCR with reverse transcription (QRT-PCR) confirmed expression of key T-lineage genes, including *Gata3*, *Bcl11b*, *CD3g*, *Lat*, *Lck* and endogenous *TCF-1* (Fig. 3b). At the time-points examined (day 10–day 14), expression of some genes in adult TCF-1-expressing *Thy1<sup>+</sup>CD25<sup>+</sup>* cells was lower than the levels in DN3 thymocytes.



**Figure 3 | TCF-1 upregulates expression of T-lineage specific genes.** **a**, Microarray-based analysis of gene expression in TCF-1-expressing *Thy1<sup>+</sup>CD25<sup>+</sup>* T cells on OP9, control *Thy1<sup>+</sup>CD25<sup>+</sup>* on OP9-DL4, and LMPPs. Shown are selected T-lineage genes upregulated greater than twofold in TCF-1-expressing T-lineage cells. Scale represents the log<sub>2</sub> value of normalized signal level. TCF-1 1 and TCF-1 2 represent biological replicates. **b**, QRT-PCR validation of selected genes normalizing to *GAPDH* and LMPP. U.D., undetectable. **c**, ChIP on DN thymocytes using TCF-1 or IgG antibodies.

QRT-PCR was performed with primers flanking putative TCF-1-binding sites. *Axin2* is a positive control and *CD3e* negative control refers to region lacking TCF-1 binding sites. **d**, ChIP as described in **c**. **e**, TCF-1 enhances TCF-1 promoter activity. 293T cells were cotransfected with the pGL3 vector containing the TCF-1 promoter and –1.3 kb TCF-1 binding site or a mutated TCF-1 binding site, and with either empty vector or MSCV-TCF-1. Luciferase activity is shown relative to Renilla and normalized to empty vector. All error bars are means  $\pm$  s.e.m. of triplicate samples. \*P < 0.05, \*\*P < 0.005.





**Figure 4 | TCF-1 is expressed in the earliest T cell progenitors and is downstream of Notch1.** **a**, LMPPs from wild-type BM were seeded onto OP9-DL4 and  $\text{Mac1}^+ \text{Gr1}^-$  cells were harvested over a 5-day period. Relative gene expression of *TCF-1* and the Notch target genes *Ptcr* and *Deltex* is shown after normalizing to 18S RNA and LMPP. **b**, *TCF-1* locus with conserved putative CSL binding sites. ChIP on DN thymocytes using Notch1 or control IgG antibodies. QRT-PCR was performed with primers flanking putative CSL-binding sites. Shown is the relative percentage of input DNA. **c**, *Scid.adh* cells were treated with 1  $\mu\text{M}$  gamma-secretase inhibitor (GSI, a pan-Notch inhibitor) or DMSO for 6 h in culture. Cells were subjected to ChIP analysis as in **b**. Shown is the relative percentage of input DNA in GSI- or DMSO-treated cultures. All error bars are means  $\pm$  s.e.m. of triplicate samples, \* $P < 0.05$ , \*\* $P < 0.005$ .

Fetal liver progenitors show accelerated differentiation *in vitro*<sup>14</sup>; TCF-1-expressing  $\text{Thy1}^+ \text{CD25}^+$  cells from fetal liver consistently expressed T-cell genes at levels comparable to DN3 thymocytes by day 10 in culture (Supplementary Fig. 11). However, some genes such as endogenous *TCF-1* and *CD3g* never reached DN3 levels, suggesting additional regulatory inputs. These data indicate ectopic expression of TCF-1 drives expression of many T-cell lineage-specific genes.

Analysis of T-lineage genes upregulated upon ectopic TCF-1 expression revealed many to contain evolutionarily conserved TCF-1 binding sites, suggesting a role for TCF-1 in directly regulating these genes. To validate these putative TCF-1 binding sites we performed chromatin immunoprecipitation assay (ChIP) on  $\text{CD4}^+ \text{CD8}^-$  (DN) thymocytes with an antibody against TCF-1. We found TCF-1 was enriched at *Gata3*, *Bcl11b*, *Il2ra*, *Cd3e* (also known as *CD3e*) and *TCF-1* itself (Fig. 3c). In addition, T-lineage genes were already upregulated in TCF-1-expressing LSK progenitors (Supplementary Fig. 12). Indeed, TCF-1 was initially cloned as a factor enriched at the *CD3e* enhancer<sup>20</sup> and TCF-1 has also been shown to regulate *Gata3* in Th2 cells<sup>21</sup>. *Gata3* is required in ETPs<sup>22</sup>, which may explain the paucity of ETPs from *TCF-1*<sup>-/-</sup> progenitors. *Bcl11b* is critical for maintenance of T-lineage commitment, as deletion of *Bcl11b* in committed T-cells results in developmental arrest or diversion to the natural killer lineage<sup>23–25</sup>.

## Regulation of TCF-1

To examine how TCF-1 expression is initially upregulated by Notch signals, we grew LMPPs on OP9-DL4. We found upregulated TCF-1 expression within 2 days that continued to rise over time, as expected<sup>8</sup> (Fig. 4a). ChIP revealed enrichment of Notch1 at a conserved -31 kilobases CSL (for CBF1, Suppressor of Hairless, and Lag-1) binding site in DN thymocytes and in 'DN3-like' *Scid.adh* cells (Fig. 4b); this binding was greatly decreased when Notch1 signals were blocked *in vitro* (Fig. 4c). The -31 kb CSL binding site was also active in a reporter assay (Supplementary Fig. 13). These data indicate Notch1 regulates TCF-1 expression.

Although TCF-1 is initially expressed downstream of Notch1 signals, TCF-1 may also regulate its own expression. TCF-1 binds to the *TCF-1* locus (Fig. 3d), and ectopic expression of human TCF-1 is sufficient to induce mouse *TCF-1* gene expression (Fig. 3b). Consistently, we found that TCF-1 activates a reporter containing the *TCF-1* promoter; mutation of the TCF-1 binding site decreased activation (Fig. 3e). Positive autoregulation may be one mechanism by which TCF-1 remains highly expressed after Notch1 signals cease after the  $\beta$ -selection checkpoint<sup>26,27</sup>, contributing to the stability of T-cell-specific gene expression.

## Conclusions

In B cells, a network of transcription factors composed of E47, EBF1, FoxO1 and Pax5 drives B-lineage gene expression<sup>28</sup>. For T-cells, similar factors were previously unknown; the present study implicates TCF-1 in this role. Our results suggest a model in which TCF-1 is induced by Notch signals in ETPs, and subsequently TCF-1 drives T-cell lineage specification. Among the genes induced by TCF-1 are components of the TCR, as well as T-cell essential transcription factors *Gata3* and *Bcl11b*. TCF-1 probably has a role in inhibiting the B-cell fate early in T-cell development, although redundant mechanisms to inhibit B-cell development from ETPs must exist<sup>29</sup>. Additional functions for TCF-1 in T-cell development and function remain to be explored in future work. The present study establishes TCF-1 as a critical regulator that is not only essential for normal T-cell development but is sufficient to establish many components of T-cell identity.

## METHODS SUMMARY

**Mice.** Mice were males or females, age 5–18 weeks. C57BL/6 (CD45.2) and B6-Ly5.2 (CD45.1) mice were purchased from the NCI animal facility. Other mice used were *Tcf7*<sup>-/-</sup> (*TCF-1*<sup>-/-</sup>  $\Delta\text{VII}$ ) mice<sup>9</sup>, *Notch1*<sup>f/f</sup> *MxCre*<sup>+</sup> *Rosa*<sup>YFP/+</sup> mice<sup>30</sup>, and  *$\beta$ -catenin*<sup>f/f</sup> *MxCre*<sup>+/-</sup> mice<sup>31</sup>. All live animal experiments were performed

according to protocols approved by the Office of Regulatory Affairs of the University of Pennsylvania in accordance with guidelines set forth by NIH.

**Intravenous transfers and intrathymic injections.** Chimaeric mice were generated by intravenously injecting T-cell-depleted *TCF-1*<sup>+/+</sup> or *TCF-1*<sup>-/-</sup> BM (CD45.2) that was mixed with wild-type T-depleted BM (CD45.1) at 1:1 or 2:1 ratios into lethally (900 rad) irradiated mice (CD45.1). Mice were analysed after 12–14 weeks for donor chimaerism. *Notch1*<sup>fl/fl</sup>*MxCre*<sup>+</sup>*Rosa*<sup>YFP/+</sup> LSK progenitors were transduced with TCF-1 or control virus; 24 h later  $2 \times 10^4$  cells were intrathymically injected into sublethally (650 rad) irradiated mice (CD45.1). Mice were analysed 10–16 days later. For intrathymic injections of TCF-1-expressing *Thy1*<sup>+</sup>*CD25*<sup>+</sup> cells, cells were isolated by cell sorting from day 8 cultures and  $3 \times 10^5$  cells were injected into sublethally irradiated mice and analysed for thymic reconstitution 1–3 weeks later.

**OP9 and OP9-DL cell culture.** OP9-GFP (OP9), OP9-DL1, and OP9-DL4 cells were provided by J. C. Zuniga-Pflucker and used as described<sup>13</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 January; accepted 9 June 2011.

- Schwarz, B. A. *et al.* Selective thymus settling regulated by cytokine and chemokine receptors. *J. Immunol.* **178**, 2008–2017 (2007).
- Spangrude, G. J. & Scollay, R. Differentiation of hematopoietic stem cells in irradiated mouse thymic lobes. Kinetics and phenotype of progeny. *J. Immunol.* **145**, 3661–3668 (1990).
- Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature Immunol.* **11**, 585–593 (2010).
- Rothenberg, E. V., Zhang, J. & Li, L. Multilayered specification of the T-cell lineage fate. *Immunol. Rev.* **238**, 150–168 (2010).
- Pui, J. C. *et al.* Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**, 299–308 (1999).
- Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of *Notch1*. *Immunity* **10**, 547–558 (1999).
- Sambandam, A. *et al.* Notch signaling controls the generation and differentiation of early T lineage progenitors. *Nature Immunol.* **6**, 663–670 (2005).
- Taghon, T. N., David, E. S., Zuniga-Pflucker, J. C. & Rothenberg, E. V. Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. *Genes Dev.* **19**, 965–978 (2005).
- Verbeek, S. *et al.* An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**, 70–74 (1995).
- Schilham, M. W. *et al.* Critical involvement of Tcf-1 in expansion of thymocytes. *J. Immunol.* **161**, 3984–3991 (1998).
- Goux, D. *et al.* Cooperating pre-T-cell receptor and TCF-1-dependent signals ensure thymocyte survival. *Blood* **106**, 1726–1733 (2005).
- Okamura, R. M. *et al.* Redundant regulation of T cell differentiation and TCR $\alpha$  gene expression by the transcription factors LEF-1 and TCF-1. *Immunity* **8**, 11–20 (1998).
- Schmitt, T. M. & Zuniga-Pflucker, J. C. T-cell development, doing it in a dish. *Immunol. Rev.* **209**, 95–102 (2006).
- Huang, J. *et al.* Propensity of adult lymphoid progenitors to progress to DN2/3 stage thymocytes with Notch receptor ligation. *J. Immunol.* **175**, 4858–4865 (2005).
- Cobas, M. *et al.*  $\beta$ -catenin is dispensable for hematopoiesis and lymphopoiesis. *J. Exp. Med.* **199**, 221–229 (2004).
- Jeannot, G. *et al.* Long-term, multilineage hematopoiesis occurs in the combined absence of  $\beta$ -catenin and  $\gamma$ -catenin. *Blood* **111**, 142–149 (2008).
- Tago, K. *et al.* Inhibition of Wnt signaling by ICAT, a novel  $\beta$ -catenin-interacting protein. *Genes Dev.* **14**, 1741–1749 (2000).
- Pear, W. S. *et al.* Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. *J. Exp. Med.* **183**, 2283–2291 (1996).
- Deftos, M. L. *et al.* Notch1 signaling promotes the maturation of CD4 and CD8 SP thymocytes. *Immunity* **13**, 73–84 (2000).
- van de Wetering, M., Oosterwegel, M., Dooijes, D. & Clevers, H. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *EMBO J.* **10**, 123–132 (1991).
- Yu, Q. *et al.* T cell factor 1 initiates the T helper type 2 fate by inducing the transcription factor GATA-3 and repressing interferon- $\gamma$ . *Nature Immunol.* **10**, 992–999 (2009).
- Hosoya, T. *et al.* GATA-3 is required for early T lineage progenitor development. *J. Exp. Med.* **206**, 2987–3000 (2009).
- Ikawa, T. *et al.* An essential developmental checkpoint for production of the T cell lineage. *Science* **329**, 93–96 (2010).
- Li, L., Leid, M. & Rothenberg, E. V. An early T cell lineage commitment checkpoint dependent on the transcription factor *Bcl11b*. *Science* **329**, 89–93 (2010).
- Li, P. *et al.* Reprogramming of T cells to natural killer-like cells upon *Bcl11b* deletion. *Science* **329**, 85–89 (2010).
- Taghon, T. *et al.* Developmental and molecular characterization of emerging  $\beta$ - and  $\gamma\delta$ -selected pre-T cells in the adult mouse thymus. *Immunity* **24**, 53–64 (2006).
- Yashiro-Ohtani, Y. *et al.* Pre-TCR signaling inactivates Notch1 transcription by antagonizing E2A. *Genes Dev.* **23**, 1665–1676 (2009).
- Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nature Immunol.* **11**, 635–643 (2010).
- Wendorff, A. A. *et al.* Hes1 is a critical but context-dependent mediator of canonical Notch signaling in lymphocyte development and transformation. *Immunity* **33**, 671–684 (2010).
- Liu, Z. *et al.* *Notch1* loss of heterozygosity causes vascular tumors and lethal hemorrhage in mice. *J. Clin. Invest.* **121**, 800–808 (2011).
- Braut, V. *et al.* Inactivation of the  $\beta$ -catenin gene by *Wnt1*-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* **128**, 1253–1264 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Clevers for permission to use *TCF-1*<sup>-/-</sup> mice and W. Pear, S. Reiner, C. Bassing, R. Sen and W. Bailis for critical comments. This work was supported by NIH grant AI059621 and a Scholar Award from the Leukemia and Lymphoma Society (A.B.), and institutional training grants T32AI055428 (B.N.W.) and T32CA09140 (A.W.-S.C.).

**Author Contributions** B.N.W., A.W.-S.C. and A.B. designed and performed the experiments. Q.Y., Y.Y.-O. and A.C. helped with experiments. O.S. provided mice and technical assistance. B.N.W. and A.B. wrote the paper.

**Author Information** Gene expression data have been deposited in the Gene Expression Omnibus database under accession numbers GSE26559 and GSE26560. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.B. ([bhandaoo@mail.med.upenn.edu](mailto:bhandaoo@mail.med.upenn.edu)).

## METHODS

**Mice.** Mice were males or females, age 5–18 weeks. C57BL/6 (CD45.2) and B6-Ly5.2 (CD45.1) mice were purchased from the NCI animal facility. Other mice used were *Tcf7<sup>-/-</sup>* (TCF-1<sup>-/-</sup> ΔVII) mice<sup>9</sup>, *Notch1<sup>fl/fl</sup>MxCre<sup>+</sup>Rosa<sup>YFP/+</sup>* mice<sup>30</sup>, and *β-catenin<sup>fl/fl</sup>MxCre<sup>+</sup>* mice<sup>31</sup>. All live animal experiments were performed according to protocols approved by the Office of Regulatory Affairs of the University of Pennsylvania in accordance with guidelines set forth by NIH.

**Intravenous transfers and intrathymic injections.** Chimaeric mice were generated by intravenously injecting T-cell-depleted TCF-1<sup>+/+</sup> or TCF-1<sup>-/-</sup> BM (CD45.2) that was mixed with wild-type T-depleted BM (CD45.1) at 1:1 or 2:1 ratios into lethally (900 rad) irradiated mice (CD45.1). Mice were analysed after 12–14 weeks for donor chimaerism. *Notch1<sup>fl/fl</sup>MxCre<sup>+</sup>Rosa<sup>YFP/+</sup>* LSK progenitors were transduced with TCF-1 or control virus; 24 h later  $2 \times 10^4$  cells were intrathymically injected into sublethally (650 rad) irradiated mice (CD45.1). Mice were analysed 10–16 days later. For intrathymic injections of TCF-1-expressing Thy1<sup>+</sup>CD25<sup>+</sup> cells, cells were isolated by cell sorting from day 8 cultures and  $3 \times 10^5$  cells were injected into sublethally irradiated mice and analysed for thymic reconstitution 1–3 weeks later.

**OP9 and OP9-DL cell culture.** OP9-GFP (OP9), OP9-DL1, and OP9-DL4 cells were provided by J. C. Zuniga-Pflucker and used as described<sup>13</sup>.

**Administration of poly(I:C).** *β-catenin<sup>fl/fl</sup>MxCre<sup>+</sup>* mice were induced as described previously<sup>32</sup>, poly(I:C) (Sigma-Aldrich) was resuspended in Dulbecco PBS at 2 mg ml<sup>-1</sup>. Mice received intraperitoneal injections of 0.2 mg poly(I:C) every other day for 2 weeks. *Notch1<sup>fl/fl</sup>MxCre<sup>+</sup>Rosa<sup>YFP/+</sup>* mice received two intraperitoneal injections of 0.2 mg poly(I:C) 1 week apart and were rested for 1 week.

**Intravenous transfers and intrathymic injections.** For intravenous transfers of transduced progenitors, wild-type LSK progenitors were transduced with TCF-1, ICN1 or control virus and transferred into sublethally irradiated mice. Mice were analysed 2–8 weeks after reconstitution for donor chimaerism in BM, spleen and thymus.

For intrathymic injection of TCF-1<sup>-/-</sup> or TCF-1<sup>+/+</sup> progenitors, fresh LSK progenitors were isolated by cell sorting and injected intrathymically. Mice were analysed after 10 days for thymic reconstitution.

**Plasmids.** MSCV-IRES-GFP (MIGR1) and MIGR1-ICN1 retroviral vectors were obtained from W. Pear. MSCV-VEX (VEX) vector was provided by C. Klug. MigR1 and VEX vectors were converted to Gateway compatible vectors and full-length TCF-1 cDNA was cloned into VEX according to the Gateway clonase manual (Invitrogen). The mouse TCF-1 promoter (~1.5 kb insert containing TCF-1 promoter activity based on Promoter Prediction 2.0, ref. 33) was cloned into pGL3 basic promoter vector. A ~1.3 kb insert containing the -31 kb CSL binding site of TCF-1 (in relation to the full-length TCF-1 translational start site) was cloned into pGL3 promoter vector (Promega). Mutation of the TCF-1 binding site in pGL3 basic-mouse TCF-1 promoter or the -31 kb CSL binding site in the pGL3 promoter vector was achieved with site-directed mutagenesis.

**Cell preparations, flow cytometry and cell sorting.** BM and thymocytes were prepared as previously described<sup>1</sup>. Cell preparations were stained with optimized antibody dilutions. Antibodies used in the lineage cocktail (Lin) include antibodies against B220 (RA3-6B2), CD19 (1D3), CD11b/Mac1 (M1/70), Gr1 (8C5), CD11c (HL3), NK1.1 (PK136), TER119 (TER-119), CD3ε (2C11), CD8α (53-6.7), CD8β (53-5.8), TCRβ (H57), γδTCR (GL-3). Additional antibodies used included antibodies against CD45<sup>B6</sup> (104), CD45<sup>SJL</sup> (A20), Sca1 (E13-161.7), Kit (2B8), Flt3 (A2F10.1), CD90.1/Thy1.1 (HIS51), Gr1 (RB6-8C5), CD19 (ID3) and CD25 (PC61.5). Antibodies were directly conjugated to fluorescein isothiocyanate (FITC), phycoerythrin (PE), PE-Cy5, PE-Cy5.5, peridinin-chlorophyll-protein complex (PerCP)-Cy5.5, PE-Cy7, allophycocyanin (APC), APC-Cy5.5 (or Alexa 700), APC-Cy7 (or APCeFluor780), or biotin. Biotinylated antibodies were revealed with Streptavidin PE-Texas Red. All antibodies were purchased from eBiosciences, Biolegend, or BD Pharmingen. Cell sorting was performed on a FACSAria II (BD Biosciences) and flow cytometric analysis was performed on a LSR-II (BD Biosciences). Dead cells were excluded through 4,6 diamidino-2-phenylindole (DAPI) uptake. Doublets were excluded through forward scatter-height by forward scatter-width and side scatter-height by side scatter-width parameters. Data were analysed using FlowJo (Tree Star). The LSK population was isolated as Lin<sup>-</sup>Sca1<sup>+</sup>Kit<sup>+</sup>. HSCs were sorted as Lin<sup>-</sup>Sca1<sup>+</sup>Kit<sup>+</sup>Flt3<sup>+</sup>CD150<sup>+</sup> BM cells; LMPPs (the 'lymphoid primed' subset of MPPs) sorted as Lin<sup>-</sup>Sca1<sup>+</sup>Kit<sup>+</sup>Flt3<sup>hi</sup> BM cells. Thymocyte populations were defined and cell-sorted as ETP (Lin<sup>-</sup>Kit<sup>+</sup>CD25<sup>-</sup>), DN2 (Lin<sup>-</sup>Kit<sup>+</sup>CD25<sup>+</sup>), DN3 (Lin<sup>-</sup>Kit<sup>+</sup>CD25<sup>+</sup>). Total thymocytes were stained and sorted as immature ISP (CD4<sup>-</sup>CD8<sup>+</sup>TCRβ<sup>-</sup>), DP (CD4<sup>+</sup>CD8<sup>+</sup>), CD4 SP (CD4<sup>+</sup>CD8<sup>-</sup>), and CD8 SP (CD8<sup>+</sup>CD4<sup>-</sup>TCRβ<sup>+</sup>).

**Retroviral transduction.** Retroviral packaging was performed as previously described<sup>5</sup>, with the exceptions of packaging cell line 293T cells and transfection

reagent FuGENE 6 (Roche) in place of CaPO<sub>4</sub>. Haematopoietic progenitors were transduced using RetroNectin (Takara). Briefly, 24 or 12-well plates were coated with 20–100 μg ml<sup>-1</sup> RetroNectin according to the manufacturer's instructions. High-titre retroviral supernatants were added into wells, centrifuged at 25 °C for 1–2 h, following which viral supernatant was removed. Cell-sorted progenitor cells were resuspended in the stimulation cocktails including DMEM-complete medium, 1% penicillin/streptomycin, 15% fetal calf serum (FCS), L-glutamate (2 mM), IL-3 (10 ng ml<sup>-1</sup>), IL-6 (10 ng ml<sup>-1</sup>), SCF (20 ng ml<sup>-1</sup>), Flt3-ligand (20 ng ml<sup>-1</sup>), Polybrene (4 μg ml<sup>-1</sup>) and added to virus-bound RetroNectin-coated plates. Transduced BM progenitors were sorted 36–48 h post-infection.

**Luciferase gene reporter assay.** For luciferase reporter assays, 293T cells were seeded 1 day before transfection to reach 80% confluency. 293T cells were transiently cotransfected with FuGENE 6 (Roche) following instructions according to manufacturer's protocol. Constructs used include: pGL3 vector (300 ng per well) containing the TCF-1 promoter with a TCF-1 binding site or a mutated TCF-1 binding site, the pGL3 promoter vector containing the wild-type -31 kb CSL binding site in TCF-1 locus or a mutated version, the TOPFLASH TCF-1 reporter, and with either empty vector MigR1, MigR1-ICAT, MigR1-TCF-1, or MigR1-ICN1 (300 ng per well). *Renilla* was added at 50 ng per well to control for transfection efficiency. DMEM containing 10% L-glutamine, 10% penicillin/streptomycin was added 24 h post transfection and cells were harvested 40–48 h after transfection and analysed with a Dual Assay Reporter Kit (Promega). Data were analysed by comparing luciferase activity to *Renilla* activity and adjusted to the fold increase over background.

**Quantitative RT-PCR.** RNA was purified from indicated cell types with the RNeasy MicroKit (Qiagen) and reverse transcribed to cDNA, using SuperScript II Kit (Invitrogen). Real-time PCR was performed with PCR Master Mix, using TaqMan probes specific for indicated genes (Applied Biosystems), and analysed on ABI Prism 7900 system (Applied Biosystems). Relative transcript abundance was determined by using the  $\Delta\Delta C_t$  or  $\Delta C_t$  method after normalization with 18S, or *GAPDH*. All samples were run in triplicate. Error bars represent s.e.m.

**ChIP.** ChIP was performed with the ChIP assay kit (Millipore), all procedures have been described<sup>27</sup>. In brief, CD4/CD8-depleted (DN) thymocytes or Scid-adh cells were fixed and immunoprecipitated with IgG control antibody (rabbit IgG; Santa Cruz Biotechnologies), Notch1 TAD/PEST-specific antiserum<sup>34</sup>, or anti-TCF-1 (C63D9) (Cell Signaling). DNA was purified using a PCR purification kit (Qiagen) and eluted by water. QRT-PCR was performed using the SYBR Green primers that flank putative TCF-1 or CSL binding sites. All genomic distances greater than 2 kb away from the translational start site were rounded to the nearest kilobase. All distances are relative to the translational start site. Primer sequences are listed in Supplementary Table 1. The relative DNA amount was calculated using the standard curve method. The input DNA was defined as an aliquot of sheared chromatin before immunoprecipitation, and was used to normalize the sample to the amount of chromatin added to each ChIP. All results are the average of triplicate PCR amplifications and results were confirmed for reproducibility in separate experiments.

**Gene expression analysis.** All protocols were conducted as described in the Affymetrix GeneChip Expression Analysis Technical Manual. RNA was extracted from sorted cells by the UPENN Microarray Core Facility, and the quality and quantity of the RNA was tested on a bioanalyser. This was followed by the Affymetrix WT Terminal Labelling kit for fragmentation and biotinylation according to the manufacturers' instructions. Biotinylated targets were heated at 99 °C for 5 min and hybridized for 16 h at 45 °C. The microarrays were then washed at low (6× SSPE) and high (100 mM MES, 0.1 M NaCl) stringency and stained with streptavidin-phycoerythrin. Fluorescence was amplified by adding biotinylated anti-streptavidin and an additional aliquot of streptavidin-phycoerythrin stain. GeneChips were scanned using the GeneArray Scanner 3000 7G. The data were analysed using Partek Genomics Suite, version 6.5 (Partek). Robust multichip average (RMA) with default settings was used to normalize data. Gene signal values for the arrays were log<sub>2</sub>-transformed and heat maps represent the log<sub>2</sub>-transformed normalized signals values or fold-change values compared to a reference population. Heat maps were generated using Matrix2png, a publicly available software<sup>35</sup>.

**Statistical analysis.** The means of each data set were analysed using Student's *t*-test, with a two-tailed distribution assuming equal sample variance.

32. Huang, J. *et al.* Pivotal role for glycogen synthase kinase-3 in hematopoietic stem cell homeostasis in mice. *J. Clin. Invest.* **119**, 3519–3529 (2009).
33. Knudsen, S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356–361 (1999).
34. Weng, A. P. *et al.* *c-Myc* is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev.* **20**, 2096–2109 (2006).
35. Pavlidis, P. & Noble, W. S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19**, 295–296 (2003).



# Forming the lunar farside highlands by accretion of a companion moon

M. Jutzi<sup>1,2</sup> & E. Asphaug<sup>1</sup>

The most striking geological feature of the Moon is the terrain and elevation dichotomy<sup>1</sup> between the hemispheres: the nearside is low and flat, dominated by volcanic maria, whereas the farside is mountainous and deeply cratered. Associated with this geological dichotomy is a compositional and thermal variation<sup>2,3</sup>, with the nearside Procellarum KREEP (potassium/rare-earth element/phosphorus) Terrane and environs interpreted as having thin, compositionally evolved crust in comparison with the massive feldspathic highlands. The lunar dichotomy may have been caused by internal effects (for example spatial variations in tidal heating<sup>4</sup>, asymmetric convective processes<sup>5</sup> or asymmetric crystallization of the magma ocean<sup>6</sup>) or external effects (such as the event that formed the South Pole/Aitken basin<sup>1</sup> or asymmetric cratering<sup>7</sup>). Here we consider its origin as a late carapace added by the accretion of a companion moon. Companion moons are a common outcome of simulations<sup>8</sup> of Moon formation from a protolunar disk resulting from a giant impact, and although most coplanar configurations are unstable<sup>9</sup>, a  $\sim 1,200$ -km-diameter moon located at one of the Trojan points could be dynamically stable for tens of millions of years after the giant impact<sup>10</sup>. Most of the Moon's magma ocean would solidify on this timescale<sup>11,12</sup>, whereas the companion moon would evolve more quickly into a crust and a solid mantle derived from similar disk material, and would presumably have little or no core. Its likely fate would be to collide with the Moon at  $\sim 2\text{--}3\text{ km s}^{-1}$ , well below the speed of sound in silicates. According to our simulations, a large moon/Moon size ratio ( $\sim 0.3$ ) and a subsonic impact velocity lead to an accretionary pile rather than a crater, contributing a hemispheric layer of extent and thickness consistent with the dimensions of the farside highlands<sup>1,13</sup> and in agreement with the degree-two crustal thickness profile<sup>4</sup>. The collision furthermore displaces the KREEP-rich layer to the opposite hemisphere, explaining the observed concentration<sup>2,3</sup>.

A study of the evolution of a terrestrial multiple-moon system in two dimensions<sup>9</sup> revealed that such systems, which are assumed to have formed from impact-generated disks, are typically unstable on short ( $<10^4$  yr) timescales owing to the rapid, tidally driven expansion of the Moon's orbit from its initial configuration very close to Earth. A moon that formed between Earth and the Moon might be stable for longer time, and stable three-dimensional configurations are plausible. A moon can also be trapped at one of the Earth–Moon Trojan points. Although the dynamics of large Trojan moons are complex, it was shown recently<sup>10</sup> that Trojans larger than 150 km in diameter can survive the Moon's orbital evolution for  $\sim 70$  Myr, at which time they pass through the weak solar resonance at 27 Earth radii. Destabilized Trojans eventually collide with the Moon or Earth at a velocity close to the mutual escape velocity.

We use a smooth particle hydrodynamics (SPH) code<sup>14,15</sup> to simulate such collisions. For the solid components of the impactor and target, we implement<sup>16</sup> a rheological model for dense granular flow<sup>17</sup>, using a block size of 5 km. For impacts, this behaves much like a Mohr–Coulomb rheology for dry friction. At global lunar scales, the hydrostatic and

impact pressure forces overwhelm the rheological stresses, such that the simulation outcomes are not very sensitive to these rheological parameters; similar results are obtained using a liquid rheology throughout. We model the Moon as a 3,500-km-diameter sphere with 700-km-diameter iron core. Thermal evolution models<sup>12</sup> indicate that  $\sim 70$  Myr after its formation, the Moon still had a small ( $\lesssim 100$ -km-deep) magma ocean. We assume that a 50-km-deep magma ocean (modelled with a liquid rheology) remains, underneath a 20-km-thick lid of crustal composition. Although these layers are thin in comparison with the impactor diameter, and do not have a significant effect on the collisional dynamics, the magma ocean is numerically resolved in our 2,500,000-particle SPH simulations (smoothing length,  $h \approx 20$  km) and the Moon's initial crust is marginally resolved. This allows us to track the dynamical evolution and redistribution of lunar material.

The companion moon, which is formed from the same protolunar disk, is similar in composition to the Moon. We model it as a 1,270-km-diameter sphere (sufficient to add 50 km of elevation to one hemisphere) consisting of lunar-like mantle below an 80-km-thick lunar-like crust. In the simulations presented here, the companion lacks an iron core; we assume that any clumps of dense iron in the protolunar disk nucleated the Moon. Simulations including a small iron core show that it sinks to join the Moon's core and does not change the overall outcome.

For the collision velocity, we assume a value of  $v_{\text{imp}} = 2.4\text{ km s}^{-1}$ , which is 1.15 times the two-body escape velocity,  $v_{\text{esc}} = [2G(M_1 + M_2)/(R_1 + R_2)]^{0.5}$ , where  $G$  is the gravitational constant and  $M_i$  and  $R_i$  ( $i = 1, 2$ ) are respectively the masses and radii of the colliding bodies. In terms of the escape velocity, our impact condition is similar to the one used in the Moon-forming giant-impact simulations (see, for example, refs 18, 19), where typically  $v_{\text{imp}} \approx v_{\text{esc}}$ . However, the different impactor/target mass ratio ( $\sim 1/10$  in refs 18, 19 versus  $\sim 1/25$  here) and, in particular, the different absolute impact velocities ( $10\text{ km s}^{-1}$  (hypervelocity) in refs 18, 19 versus  $2.4\text{ km s}^{-1}$  (subsonic) here) indicate that these are two different types of collisions. This can also be illustrated by computing the specific incoming kinetic energy,  $Q = (1/2)M_2 v_{\text{imp}}^2/M_1$ , which is  $\sim 40$  times larger in the Moon-forming giant impact than in our proposed Moon–moon collision. (Here  $M_2$  is the mass of the impactor and  $M_1$  is the mass of the target.)

We compute self-gravity using a grid-based gravity solver. The final thermal state is important to the long-term fate of the accreted material, but because the collision is subsonic, the post-impact thermal state is for the most part advected from the pre-impact state. Impact melting is not important at these velocities, but the presence of melt within the Moon is expected during the first  $\sim 10\text{--}100$  Myr (ref. 11). Both the target and the impactor are evolved to hydrostatic equilibrium in the simulation before the collision. We use the Tillotson equation of state for iron (the Moon's core) and dunite (the Moon's mantle and crust and the companion moon) (compare with ref. 20). The initial densities of the target materials are  $7.8\text{ g cm}^{-3}$  (core),  $3.4\text{ g cm}^{-3}$  (mantle and magma ocean) and  $2.9\text{ g cm}^{-3}$  (crust). For the impactor, we use  $3.0\text{ g cm}^{-3}$  (mantle) and  $2.9\text{ g cm}^{-3}$  (crust). We consider two impact angles,  $0^\circ$  (head-on collision) and  $45^\circ$ . Because of the initial, presumably tidally locked state,

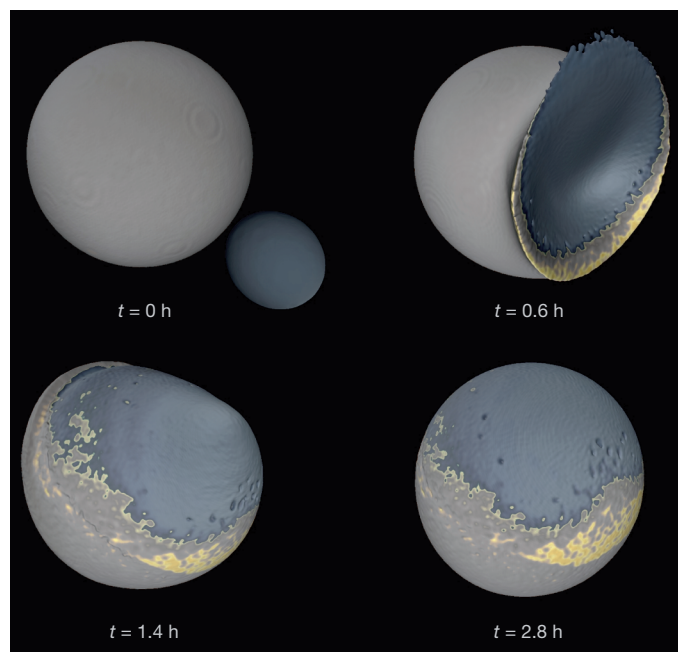
<sup>1</sup>Earth and Planetary Sciences Department, University of California, Santa Cruz, 1156 Highstreet, Santa Cruz, California 95060, USA. <sup>2</sup>Physics Institute, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.

we assume zero pre-impact rotation. We have not considered post-impact diffusion by short-term (viscous flow or seismic fluidization) or long-term (tidal/thermal) processes.

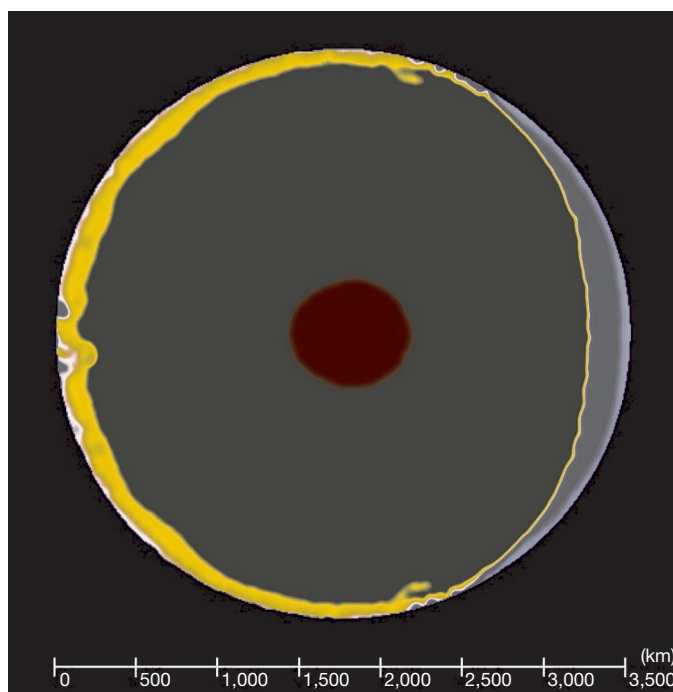
Our primary finding is that a companion moon  $\sim 1/3$  the diameter of the Moon, striking at subsonic velocity, does not form a crater. For this low impact velocity and large impactor scale, the impactor volume,  $V_{\text{imp}}$ , exceeds the excavated volume,  $V_{\text{exc}}$ , and the impact produces an accretionary pile rather than a crater. Although our initial conditions are clearly not those that are assumed by point-source scaling laws<sup>21</sup>, it is interesting to note that, assuming a flat geometry, the scaling laws predict that  $V_{\text{imp}}/V_{\text{exc}} \approx 5$ , that is, much more added material than excavated volume.

Most of the colliding material stays local to the impact (Fig. 1), pasting on a thickened crust and forming a mountainous region comparable in extent to the farside highlands. Underneath the accreted impactor, the Moon's magma ocean is displaced to the opposite hemisphere, as can be seen in cross-section in Fig. 2. As can also be seen in Figs 1 and 2, the initial crust of the impactor stays on top. According to our model, the post-impact density gradient of the hemisphere with the accreted impactor is stable.

Our choice of initial densities is based on the assumption that the impactor accreted from materials with densities no greater than that of the Moon. At  $\sim 1/10$  the interior pressure of the Moon, the moon's solids would have been less compacted; the moon's mass density would also be lower if the Moon accreted around any and all iron in the disk. If the impacting moon was for some reason more dense than the Moon, then the collisional outcome and distribution of material would be about the same, but the resulting structure might be unstable in the long term, resulting in a denser hemispheric pancake of material that might



**Figure 1 | Moon/companion moon collision.** We use SPH to simulate collisions between the Moon and a companion moon,  $\sim 4\%$  the lunar mass, dislodged from one of the Earth–Moon Trojan points, to explore whether the Moon's farside highlands and its nearside KREEP-rich terrain can be explained by this late, slow accretion. Snapshots ( $t$ , simulation time) show the case of a  $2.4 \text{ km s}^{-1}$ ,  $45^\circ$  impact of a  $1,270\text{-km}$ -diameter Trojan moon impacting the  $3,500\text{-km}$ -diameter Moon. Plotted is an iso-density surface  $\rho_{\text{iso}} = 2 \text{ g cm}^{-3}$ ; lower density materials are invisible. Plotted colours indicate impactor crust (light blue), impactor mantle (dark blue), target crust (grey) and a layer of target upper mantle material (yellow) representing a magma ocean. Most of the impactor is accreted as a pancake-shaped layer, forming a mountainous region comparable in extent to the lunar farside highlands. A residual magma ocean, if present, gets pushed (Fig. 2) to the opposite hemisphere.



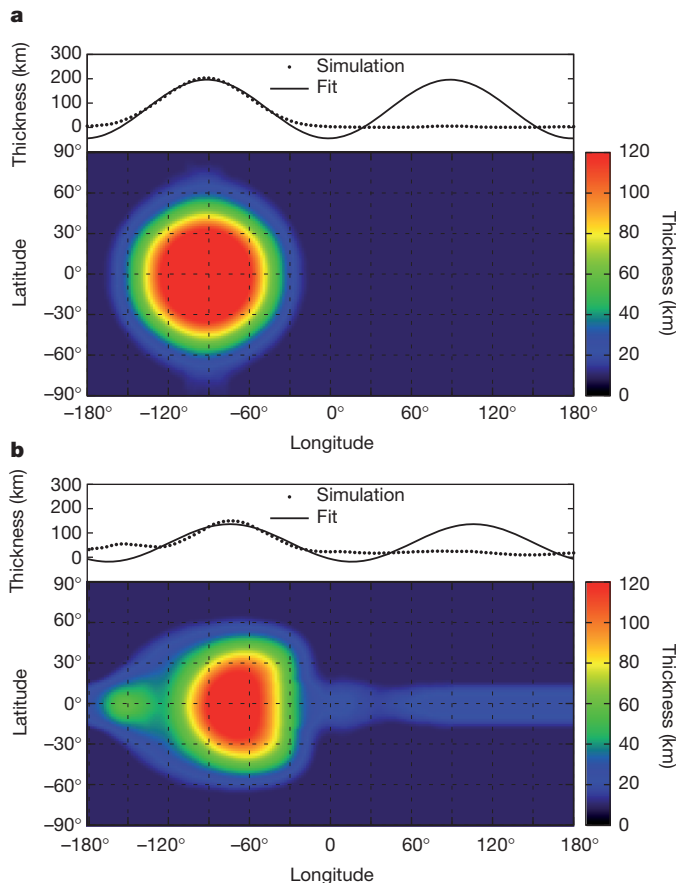
**Figure 2 | Post-impact internal structure.** Two-dimensional view of the post-impact distribution of layers of target and impactor materials for the case of the head-on impact simulation. The companion moon collided from the right, accreting as a pile and producing the farside lunar highlands in our model. Grey and light blue correspond to the mantle and crust of the companion moon, respectively. The initially global residual magma ocean (yellow) is displaced to the opposite hemisphere, leading to an asymmetric distribution of KREEP. The initial  $\sim 20\text{-km}$ -thick crust of the target (white) is poorly resolved in the code, and not shown to scale.

founder. Although this could provide another explanation of hemispheric dichotomy, we consider the more reasonable picture of a companion moon no denser than the Moon, and consequently stable layering.

The impacting moon, forming in the same giant impact event as the Moon but solidifying more rapidly in a smaller sized body, has an older crust. Our model therefore predicts two differing ages for lunar anorthositic crustal rocks. A fraction of the older crust (from the impactor) ends up on the nearside in the  $45^\circ$  impact simulation (Fig. 3), and we expect further widespread redistribution of crustal material following subsequent basin-forming events. Measured lunar crustal ages do show a wide spread<sup>11</sup> ( $\sim 200$  Myr), which is consistent with our picture of two compositionally similar crustal reservoirs.

In Fig. 3, we plot the post-impact thickness of the accreted moon material in a cylindrical projection of the Moon. For both impact angles considered here, the resulting pancake of material is comparable in extent to the farside highlands. A profile of the resulting crustal thickness for latitude  $0^\circ$  is also shown in Fig. 3. In the impacted hemisphere, our result can be very well fitted by a degree-two Legendre polynomial. Although our model reproduces the observed degree-two terrain well, other geological mechanisms can in principle also lead to similar terrain forms. For instance, the degree-two terrain has been interpreted<sup>4</sup> as evidence for spatial variation in tidal heating. The nearside would in that case require a second major event or a process of crustal thinning. In our model, the accretion of a large impactor is shown to result in a highlands region with a degree-two thickness structure, in accordance with the observations, and only on the impacted hemisphere.

Our resulting degree-two terrain profiles have amplitudes greater than the observed variations in crustal thickness<sup>1,4,13</sup>. A comprehensive analysis of our simulation results would use the density structure of the



**Figure 3 | Post-impact spatial distribution of the impactor and thickness profile.** Thickness of the impactor layer accreted onto the Moon, shown as a function of longitude ( $-180^{\circ}$  to  $180^{\circ}$ ) and latitude ( $-90^{\circ}$  to  $90^{\circ}$ ), after a 1,270-km-diameter moon impacts at  $2.4 \text{ km s}^{-1}$ : head-on impact (a),  $45^{\circ}$  impact from the right (b). The 'bull's-eye' would be the lunar farside. Also shown are a thickness profile computed for the great circle at latitude  $0^{\circ}$  and, for comparison, a degree-two Legendre polynomial that was fitted to the thickness profile.

accreted impactor on the Moon to compute a modelled post-impact Bouguer gravity anomaly, for direct comparison with observations<sup>1</sup>. Such a detailed gravity and terrain comparison is complicated because we must assume a pre-impact structure for the Moon and to be accurate must resolve the Moon's crust at the time of impact using three to five particles, requiring  $\sim 10^8$  particle simulations and much faster computation. Perhaps more importantly, the post-impact shape may not be well preserved if the Moon is still significantly molten.

KREEP is found concentrated in the Procellarum KREEP Terrane, one of the three major geologic provinces of the Moon<sup>3</sup>, situated on the nearside. According to the magma ocean model<sup>22</sup>, the incompatible elements of which it is composed (potassium, phosphorus and the rare-earth elements) are concentrated in the last magmas to crystallize and are sandwiched between the floating anorthositic crust and the mantle cumulates. In some models<sup>23</sup>, the Moon's KREEP asymmetry is explained as a result of a degree-one Rayleigh–Taylor instability. Our model provides an alternative explanation. This is illustrated in Fig. 2, where the magma ocean at the time of the impact (the global liquid layer) is pushed/ejected to the opposite hemisphere in response to the same forces that flatten and accrete the impactor.

One obvious test of our model would be to determine whether there is evidence of accretion of a foreign-composition body on the Moon's farside. However, the companion moon formed from the same protolunar disk as the Moon, evolved similarly (although faster, being smaller) and was not shocked by the final collision, so we expect compositional distinctions to be subtle. A more important test will

be to compare the lunar farside with the high-definition gravimetric determinations of the crust and mantle in the upcoming data sets from the NASA Gravity Recovery and Interior Laboratory mission<sup>24</sup>, and others, and perhaps more clearly in forthcoming lunar seismology experiments. In particular, if the KREEP was actually pushed from one hemisphere to the other, this deep-scale evolution should be preserved. Finally, we should acknowledge that although the dynamical modelling<sup>10</sup> indicates the possibility of a Trojan moon surviving for tens of millions of years, there is much work to be done studying the long-term after-effects of Moon-forming giant impacts.

## METHODS SUMMARY

We use an SPH impact code<sup>14,15</sup> specially written to model geologic materials. Our code computes the full stress tensor,  $\sigma_{ij} = -P\delta_{ij} + \tau_{ij}$  where  $P$  denotes pressure,  $\delta_{ij}$  is the Kronecker delta function and  $\tau_{ij}$  is the deviatoric stress tensor. For solids, we apply a model for dense granular flow<sup>17</sup> that has a Drucker–Prager-like yield criterion: material flows only if  $|\tau| > \mu_s P$ , where  $|\tau|$  is the second invariant of  $\tau_{ij}$  and  $\mu_s = 0.7$  is the static friction coefficient. A yield stress of  $P$  or 0.1 GPa, whichever is the lesser, is applied to the computation of  $\sigma_{ij}$ , and tensile strength is neglected in these simulations. Comparison simulations using a liquid rheology showed no significant differences in terms of the final distribution of impactor material. Self-gravity is computed using a grid-based solver. Both the target and the impactor are evolved to hydrostatic equilibrium before the collision. The initial separation between the impactor and the target is one target radius; the impactor is assigned an initial velocity of  $1.8 \text{ km s}^{-1}$ , which leads to an impact velocity of  $2.4 \text{ km s}^{-1}$ . The collision simulations are carried out over a simulation time of  $\sim 10 \text{ h}$ , which takes about seven days on 32 computer processors for a resolution of  $2.5 \times 10^6$  SPH particles.

Received 8 April; accepted 13 June 2011.

1. Zuber, M. T., Smith, D. E., Lemoine, F. G. & Neumann, G. A. The shape and internal structure of the moon from the Clementine mission. *Science* **266**, 1839–1843 (1994).
2. Lawrence, D. J. Global elemental maps of the moon: the Lunar Prospector gamma-ray spectrometer. *Science* **281**, 1484–1489 (1998).
3. Jolliffe, B. L., Gillis, J. J., Haskin, L. A., Korotev, R. L. & Wieczorek, M. A. Major lunar crustal terranes: surface expressions and crust-mantle origins. *J. Geophys. Res.* **105**, 4197–4216 (2000).
4. Garrick-Bethell, I., Nimmo, F. & Wieczorek, M. Structure and formation of the lunar farside highlands. *Science* **330**, 949–951 (2010).
5. Loper, D. E. & Werner, C. L. On lunar asymmetries 1. Tilted convection and crustal asymmetry. *J. Geophys. Res. Planets* **107**, 131–137 (2002).
6. Wasson, J. T. & Warren, P. H. Contribution of the mantle to the lunar asymmetry. *Icarus* **44**, 752–771 (1980).
7. Wood, J. A. Bombardment as a cause of the lunar asymmetry. *Moon* **8**, 73–103 (1973).
8. Ida, S., Canup, R. M. & Stewart, G. R. Lunar accretion from an impact-generated disk. *Nature* **389**, 353–357 (1997).
9. Canup, R. M., Levison, H. F. & Stewart, G. R. Evolution of a terrestrial multiple-moon system. *Astron. J.* **117**, 603–620 (1999).
10. Cuk, M. & Gladman, B. J. The fate of primordial lunar Trojans. *Icarus* **199**, 237–244 (2009).
11. Elkins-Tanton, L. T., Burgess, S. & Yin, Q.-Z. The lunar magma ocean: reconciling the solidification process with lunar petrology and geochronology. *Earth Planet. Sci. Lett.* **304**, 326–336 (2011).
12. Meyer, J., Elkins-Tanton, L. & Wisdom, J. Coupled thermal–orbital evolution of the early moon. *Icarus* **208**, 1–10 (2010).
13. Wieczorek, M. The constitution and structure of the lunar interior. *Rev. Mineral. Geochem.* **60**, 221–364 (2006).
14. Benz, W. & Asphaug, E. Simulations of brittle solids using smooth particle hydrodynamics. *Comput. Phys. Commun.* **87**, 253–265 (1995).
15. Jutzi, M., Benz, W. & Michel, P. Numerical simulations of impacts involving porous bodies. I. Implementing sub-resolution porosity in a 3D SPH hydrocode. *Icarus* **198**, 242–255 (2008).
16. Jutzi, M. & Asphaug, E. Mega-ejecta on asteroid Vesta. *Geophys. Res. Lett.* **38**, L01102 (2011).
17. Jop, P., Forterre, Y. & Pouliquen, O. A constitutive law for dense granular flows. *Nature* **441**, 727–730 (2006).
18. Benz, W. The origin of the moon and the single-impact hypothesis. I. *Icarus* **66**, 515–535 (1986).
19. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).
20. Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, 1989).
21. Housen, K. R., Schmidt, R. M. & Holsapple, K. A. Crater ejecta scaling laws: fundamental forms based on dimensional analysis. *J. Geophys. Res.* **88**, 2485–2499 (1983).
22. Warren, P. H. The magma ocean concept and lunar evolution. *Annu. Rev. Earth Planet. Sci.* **13**, 201–240 (1985).



23. Parmentier, E. M., Zhong, S. & Zuber, M. T. Gravitational differentiation due to initial chemical stratification: origin of lunar asymmetry by the creep of dense KREEP? *Earth Planet. Sci. Lett.* **201**, 473–480 (2002).
24. Zuber, M. T. *et al.* in *Proc. 39th Lunar Planet. Sci. Conf.* abstr. 1074, (<http://www.lpi.usra.edu/meetings/lpsc2008/pdf/1074.pdf>) (Lunar and Planetary Institute, 2008).

**Acknowledgements** Our work is sponsored by NASA's Planetary Geology and Geophysics programme 'Small Bodies and Planetary Collisions'. All simulations were run at the University of California, Santa Cruz, on the NSF-MRI-sponsored

'pleiades' cluster. We are grateful to M. Cuk, B. Gladman and R. Canup for guidance.

**Author Contributions** Both authors contributed equally to this work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.J. ([martin.jutzi@space.unibe.ch](mailto:martin.jutzi@space.unibe.ch)).

# Link between spin fluctuations and electron pairing in copper oxide superconductors

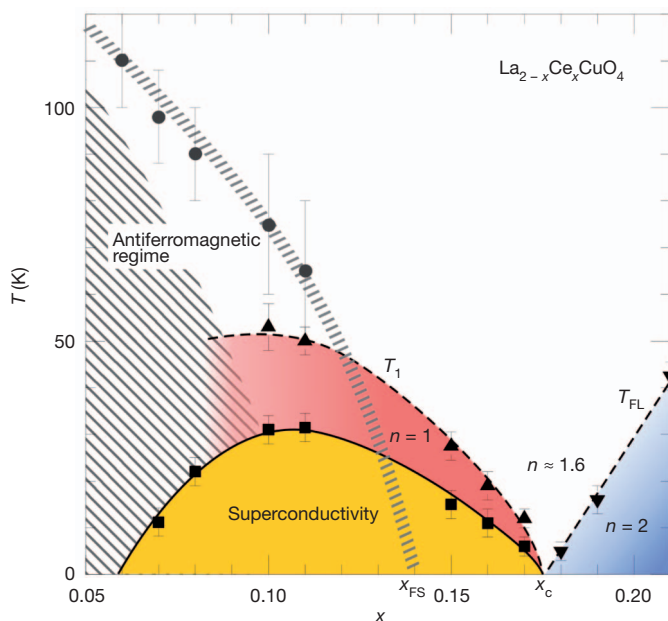
K. Jin<sup>1,2</sup>, N. P. Butch<sup>1,2</sup>, K. Kirshenbaum<sup>1,2</sup>, J. Paglione<sup>1,2</sup> & R. L. Greene<sup>1,2</sup>

Although it is generally accepted that superconductivity is unconventional in the high-transition-temperature copper oxides, the relative importance of phenomena such as spin and charge (stripe) order, superconductivity fluctuations, proximity to a Mott insulator, a pseudogap phase and quantum criticality are still a matter of debate<sup>1</sup>. In electron-doped copper oxides, the absence of an anomalous pseudogap phase in the underdoped region of the phase diagram<sup>2</sup> and weaker electron correlations<sup>3,4</sup> suggest that Mott physics and other unidentified competing orders are less relevant and that antiferromagnetic spin fluctuations are the dominant feature. Here we report a study of magnetotransport in thin films of the electron-doped copper oxide  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$ . We show that a scattering rate that is linearly dependent on temperature—a key feature of the anomalous normal state properties of the copper oxides—is correlated with the electron pairing. We also show that an envelope of such scattering surrounds the superconducting phase, surviving to zero temperature when superconductivity is suppressed by magnetic fields. Comparison with similar behaviour found in organic superconductors<sup>5</sup> strongly suggests that the linear dependence on temperature of the resistivity in the electron-doped copper oxides is caused by spin-fluctuation scattering.

Resistivity that increases linearly with temperature (linear-temperature resistivity) is well known to appear in proximity to an antiferromagnetic quantum critical point (QCP), as found in organic<sup>5</sup> and heavy-fermion<sup>6</sup> strongly correlated materials. Unlike the hole-doped copper oxides, the absence of anomalous pseudogap physics and other unidentified competing phases in these materials allows such non-Fermi-liquid properties to be attributed to the presence of an antiferromagnetic QCP (ref. 6). This has led to models that ascribe linear-temperature resistivity to a mechanism involving spin fluctuation scattering<sup>7–9</sup>. The case for this is particularly strong in the Bechgaard class of organic superconductors  $(\text{TMTSF})_2\text{PF}_6$ , where scattering that increases linearly with temperature (linear-temperature scattering) dominates the normal-state transport above a superconducting state induced by the suppression of a spin-density-wave order by applied pressure<sup>5</sup>. The anisotropic two-dimensional nature of the Bechgaard compounds allows for microscopic calculations of the interdependence of antiferromagnetic and superconducting correlations<sup>10</sup>, yielding a thorough understanding of the origin of the anomalous scattering rate in this case<sup>5,11</sup>. However, in general, no microscopic theory yet exists for the origin of linear-temperature scattering at low temperatures. In  $(\text{TMTSF})_2\text{PF}_6$ , the linear-temperature scattering rate found at the spin-density-wave QCP has been shown to be suppressed with pressure along with the superconducting transition, with a scattering coefficient that approaches zero along with the transition temperature  $T_c$  (refs 5, 11). In electron-doped  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$  (PCCO), linear-temperature resistivity is found down to 35 mK at  $x = 0.17$  (ref. 12). Along with other evidence for a Fermi-surface reconstruction<sup>2,13–15</sup>, this observation suggests that an antiferromagnetic QCP occurs near  $x = 0.17$  in PCCO.

$\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$  (LCCO) is an electron-doped copper oxide<sup>16</sup> with properties very similar to PCCO, but with a superconductivity dome

that is slightly shifted towards lower Ce concentrations such that the superconducting phase exists for  $0.06 \leq x \leq 0.17$  and is suppressed for  $x > 0.17$ . The phase diagram of LCCO (Fig. 1), constructed from our



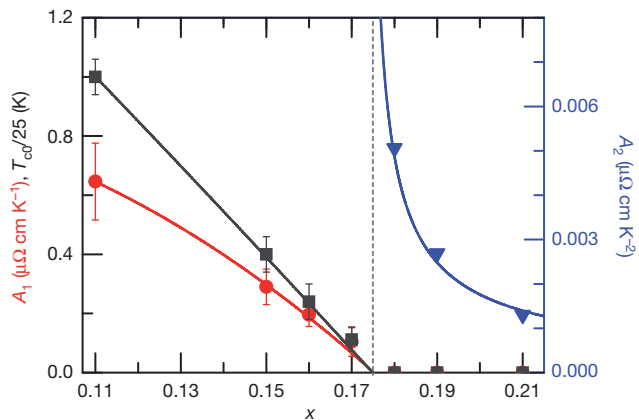
**Figure 1 | Temperature-doping ( $T$ - $x$ ) phase diagram of  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$ .**

The resistivity in zero field can be expressed by  $\rho = \rho_0 + AT^n$ , with  $n = 1$  and  $2$  for the red and blue regimes, respectively. Between the  $\rho \propto T$  ( $n = 1$ ) and the Fermi-liquid ( $n = 2$ ) regimes, the data below 50 K is well fitted by a single power law with  $n \approx 1.6$ . The yellow regime is the superconductivity dome. The superconductivity,  $\rho \propto T$  and Fermi-liquid regimes terminate at one critical doping,  $x_c$ . The temperatures  $T_1$  (triangles) and  $T_{FL}$  (inverted triangles) mark the crossover temperatures to the  $\rho \propto T$  and Fermi-liquid regimes, respectively. To illustrate the  $\rho \propto T$  regime more clearly, the boundary of the superconductivity dome (squares) is defined as the lowest temperature of the linear-temperature resistivity for  $x \geq 0.1$ . For  $x < 0.1$ , the resistivity shows an upturn (hatched area) with decreasing temperature, a typical feature of underdoped copper oxides. Owing to the upturn, the superconductivity boundary for  $x < 0.1$  is defined as the temperature where the resistivity reaches zero ( $T_{c0}$ ). The antiferromagnetic (or spin-density-wave) regime (circles) is estimated from previous in-plane angular magnetoresistance measurements<sup>18</sup>. A QCP associated with a spin-density-wave Fermi surface reconstruction is estimated to occur near  $x = 0.14$  (indicated as  $x_{FS}$ ). LCCO can only be prepared in thin-film form, so the evidence for a spin-density-wave (antiferromagnetic) QCP under the superconductivity dome is not as conclusive as for the electron-doped copper oxides  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$  or  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ . Nevertheless, in LCCO the change of the sign of the low-temperature Hall coefficient at  $x \approx 0.14$  (ref. 19), angular magnetoresistance data, and a low-temperature metal-to-insulator crossover at  $x \approx 0.14$  (ref. 16) all suggest that such a QCP, associated with Fermi surface reconstruction, does occur near  $x = 0.14$ . The error bars on the circles are from ref. 18 and those on other symbols represent the standard error in the fit to the data.

<sup>1</sup>Center for Nanophysics & Advanced Materials, University of Maryland, College Park, Maryland 20742, USA. <sup>2</sup>Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

present transport studies on optimal to overdoped thin films ( $x \geq 0.11$ ) and prior work<sup>17–19</sup> for  $x < 0.12$ , has four distinct regions: the superconducting phase, the linear-temperature ( $\rho \propto T$ ) region, the non-Fermi liquid ( $\rho \propto T^{1.6}$ ) region and the Fermi-liquid ( $\rho \propto T^2$ ) region. In the superconductivity doping range, all films exhibit a linear-temperature resistivity above  $T_c$  that extends from  $T_c$  up to a characteristic crossover temperature  $T_1$ , forming a shell of anomalous scattering that encases the superconductivity dome. For example, the resistivity of optimally doped  $x = 0.11$  is linear from  $T_c$  up to  $T_1 \approx 45$  K (Supplementary Fig. 1). For higher doping, this temperature range (and thus  $T_1$ ) decreases, tending towards zero along with  $T_c$  itself at the end of the superconductivity dome at a critical doping of  $x_c = 0.175 \pm 0.005$ . In PCCO, a similar phenomenon is observed (Supplementary Figs 2 and 3) with a linear-temperature region above  $T_c$  that extends as far as films can be synthesized (that is, up to  $x = 0.19$ ). Similarly, in hole-doped LSCO ( $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ ) a linear-temperature component was shown to diminish upon approaching the end of the doping range of superconductivity<sup>20</sup>, suggestive of a common relation between scattering and pairing in both electron- and hole-doped copper oxides. The nature of the QCP in hole-doped copper oxides remains uncertain. Note, however, that a linear resistivity identical to that of LSCO (ref. 20) was observed in  $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$  (Nd-LSCO) (ref. 21) at the QCP, where stripe order is known to end.

A direct relation between linear-temperature scattering and  $T_c$  is revealed through the doping dependence of each. As shown in Fig. 2, the scattering coefficient  $A_1(x)$ , obtained from fits to the linear-temperature regions with  $\rho(T) = \rho_0 + A_1(x)T$ , decreases with  $T_c$  as  $x$  is increased and approaches zero at the critical doping  $x_c$ . This scaling of  $A_1$  with  $T_c$  is also observed in PCCO (Supplementary Fig. 4), indicating that it is not specific to the doping concentration (which is

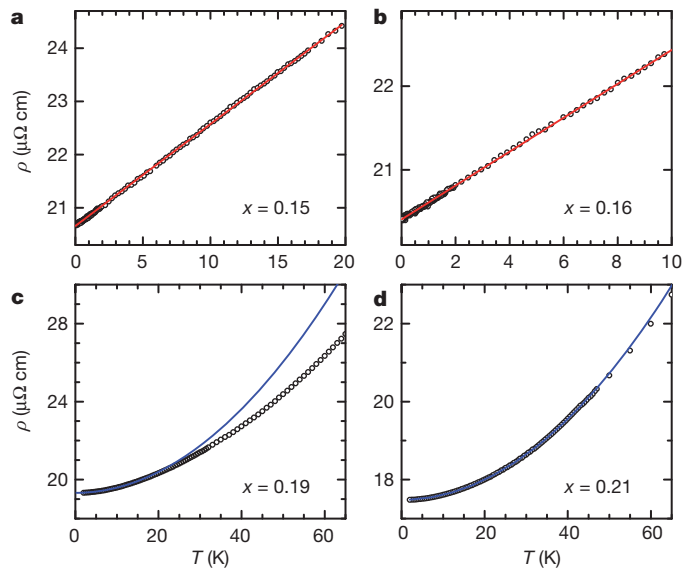


**Figure 2 | Doping dependence of scattering rates in zero field.** The left vertical axis shows linear-temperature scattering rate  $A_1$  (red circles) and also  $T_{c0}$  (divided by 25; black squares) versus  $x$ . The right vertical axis shows quadratic scattering rate  $A_2$  (blue triangles) versus  $x$ . For the superconducting LCCO films with  $x < 0.18$ ,  $A_1$  data are obtained from the  $\rho \propto T$  region ( $\rho = \rho_0 + A_1 T$ , the red regime in Fig. 1). The error bars are the standard deviation over many samples of each doping. We note that in the optimally doped region, the highest superconducting transition temperatures of  $x = 0.1$  and  $0.11$  are almost the same by a slight oxygen variation, and their resistivity also shows similar behaviour. Thus, only one nominal  $x = 0.1$  sample was studied here; nevertheless, both the  $A_1$  and  $T_{c0}$  (data not shown) fall into the statistical error of the  $x = 0.11$  samples. We use the  $x = 0.11$  doping to represent the optimal doping level here. For the non-superconducting films with  $x \geq 0.18$ ,  $A_2$  data are obtained from the  $\rho \propto T^2$  region ( $\rho = \rho_0 + A_2 T^2$ , the blue regime in Fig. 1). It is noteworthy that the amplitude of the linear-temperature scattering scales with the superconductivity transition temperature (both ending around  $x_c = 0.175$ ), reflecting the intimate relationship between the linear-temperature scattering rate and the superconductivity. From the non-superconductivity side, as the doping approaches  $x_c$  from higher doping, the coefficient of electron–electron scattering increases very quickly, reminiscent of critical scattering upon approach to a QCP.

shifted in PCCO compared to LCCO for a given  $T_c$ ), but is representative of a central relationship between  $T_c$  and  $A_1$ . The same relation has been found in  $(\text{TMTSF})_2\text{PF}_6$  (refs 5, 11), reflecting the intimate connection between the strength of the linear-temperature inelastic scattering and the electron pairing in systems governed by spin fluctuations. Similar scaling is seen in the hole-doped copper oxides LSCO, Nd-LSCO and  $\text{Ti}_2\text{Ba}_2\text{CuO}_{6+\delta}$  (ref. 11), again suggesting that the physics of scattering and pairing is the same in electron- and hole-doped copper oxides.

The linear-temperature scattering is robust and survives in magnetic fields exceeding the upper critical field for superconductivity of LCCO. In fact, when superconductivity is completely suppressed, the linear-temperature resistivity extends down to the  $T = 0$  limit without any indication of saturation or change in behaviour. For instance, for  $x = 0.15$  at 7.5 T (Fig. 3a, Supplementary Fig. 1), linear-temperature resistivity extends from  $T \approx 20$  K down to the lowest measured temperature of 20 mK. Spanning over three decades in temperature, this behaviour clearly points to a scattering mechanism that originates from an anomalous ground state. Similar behaviour is found at higher  $x$  (Fig. 3b), but occurs over a decreasing range as  $T_c$  is suppressed to zero with doping, again suggesting that linear-temperature scattering is intimately tied to the presence of superconductivity.

Many experiments have shown that spin fluctuations dominate the physical properties in proximity to a critical doping under the superconductivity dome in the more-studied electron-doped copper oxides PCCO and  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$  (refs 2, 22, 23). In analogy with these other electron-doped copper oxides, it is expected that the boundary of antiferromagnetic order in LCCO extrapolates to a QCP beneath the superconductivity dome (indicated as  $x_{\text{FS}}$  in Fig. 1, where subscript ‘FS’ indicates Fermi surface), having a fundamental role in generating the superconducting phase. In particular, the extended linear-temperature transport scattering that persists to the lowest measurable temperatures is exactly in line with that expected at an antiferromagnetic QCP



**Figure 3 | Temperature dependence of normal-state resistivity.** a and b,  $\rho(T)$  of  $x = 0.15$  and  $0.16$  LCCO films in a perpendicular magnetic field where the superconductivity is just suppressed, that is, at 7.5 and 7 T, respectively. The data can be fitted by  $\rho = \rho_0 + A_1 T$  down to the lowest measuring temperature. The linearity of the resistivity of  $x = 0.15$  persists from 20 K down to 20 mK, spanning over three decades in temperature. That is, the  $\rho \propto T$  region shown in Fig. 1 can extend down to the  $T = 0$  limit, pointing to a scattering mechanism that originates from an anomalous ground state. c and d,  $\rho(T)$  of  $x = 0.19$  and  $0.21$  in zero field, fitted by  $\rho = \rho_0 + A_2 T^2$  (blue lines). In the non-superconductivity regime ( $x \geq 0.18$ ), the Fermi-liquid behaviour can also persist to the lowest temperature, that is, down to 20 mK (as seen in Supplementary Fig. 5 for  $x = 0.18$ ).



for a two-dimensional disordered Fermi-liquid system<sup>9</sup>. Moreover, inelastic neutron scattering experiments on electron-doped  $\text{Pr}_{1-x}\text{LaCe}_x\text{CuO}_{4+\delta}$  show that the strength of the spin fluctuations decreases with overdoping in the superconducting phase and that these fluctuations disappear at the end of the superconductivity dome<sup>24</sup>.

Non-superconducting films of LCCO doped beyond  $x_c$  exhibit a  $T^2$  dependence of  $\rho(T)$  in the low-temperature limit, indicating a conventional Fermi-liquid behaviour due to electron-electron scattering, similar to that exhibited by  $(\text{TMTSF})_2\text{PF}_6$  (ref. 5) and LSCO (ref. 25). For example, LCCO films with  $x = 0.18$  exhibit a  $T^2$  resistivity up to 5 K, spanning over two orders of magnitude in temperature (Supplementary Fig. 5). The highest temperature of the quadratic behaviour ( $T_{\text{FL}}$ , where subscript FL refers to Fermi liquid) increases with increasing  $x$  as shown for  $x = 0.19$  and  $0.21$  (Fig. 3c and d), and notably, this line extrapolates to  $T = 0$  at  $x_c$ .

In LCCO, the critical doping  $x_c$  is exactly where the superconductivity dome terminates and the two characteristic crossover temperatures  $T_1$  and  $T_{\text{FL}}$  approach absolute zero. Interestingly, indications of the singular nature of  $x_c$  are evident even from within the overdoped Fermi liquid regime of LCCO. In this region of the phase diagram, the coefficient of electron-electron scattering  $A_2(x)$  (that is, obtained from fits to  $\rho(T) = \rho_0 + A_2(x)T^2$ ) exhibits a strong enhancement upon approach to  $x_c$  from higher doping, reminiscent of critical scattering upon approach to a QCP (ref. 6). This suggests that the onset of superconductivity marks a dramatic change in the ground state and its excitations. While Fermi-liquid behaviour of resistivity has been reported at one doping in both hole-doped LSCO (ref. 25) and  $\text{Ti}_2\text{Ba}_2\text{CuO}_{6+\delta}$  (ref. 26), such doping-tuned critical behaviour in the non-superconducting region was not observed. In LCCO, the resistivity directly above the critical point at  $x = 0.175$  and in the entire temperature regime above the characteristic temperatures  $T_1$  and  $T_{\text{FL}}$  is best fitted by a single power-law dependence,  $\rho = \rho_0 + A'T^n$  with  $n \approx 1.6$ , up to at least 50 K (Supplementary Fig. 6). Perhaps not coincidentally, the same power law is observed above the Fermi-liquid ( $\propto T^2$ ) regime in LSCO (refs 20, 25), signifying that scattering throughout the non-Fermi-liquid regime is governed by the same physics in both hole- and electron-doped copper oxides. Clearly, our observation of critical behaviour at  $x_c$  will require further experimental and theoretical investigation to determine its significance for the unusual transport properties of the copper oxides.

With the absence of anomalous pseudogap phenomena in electron-doped copper oxides<sup>2</sup>, comparisons to similarly tractable systems allow for far-reaching conclusions to be drawn. Studies<sup>5,10,11</sup> of the organic superconductor  $(\text{TMTSF})_2\text{PF}_6$  show that electron pairing and linear-temperature scattering arise from antiferromagnetic (spin-density-wave) spin fluctuations. Given the very similar experimental transport properties and evolution of ground states in the phase diagram of LCCO, it is likely that the scattering and pairing in the electron-doped copper oxides is governed by a similar interplay of spin fluctuations and superconductivity. The results of our work reported here, and their analogy to  $(\text{TMTSF})_2\text{PF}_6$ , strongly suggests that the pairing in electron-doped copper oxides is not coming from phonons or any other unusual pseudogap order parameter (such as  $d$ -density waves, orbital currents or stripe order), but rather from spin-fluctuation-mediated pairing<sup>27–29</sup>. The striking similarities between transport properties of electron- and hole-doped copper oxides provides evidence that the mechanism of the anomalous linear-temperature scattering rate and high- $T_c$  pairing are shared between the two families, and, furthermore, bear a striking resemblance to simpler systems well described by the spin fluctuation scenario. Although the role of the pseudogap and unidentified competing phases in the hole-doped copper oxides remains to be conclusively determined, the similar correlation between the linear-temperature scattering and  $T_c$  for both electron- and hole-doped copper oxides suggests that spin fluctuations also play a crucial part in hole-doped copper oxides.

Received 4 March; accepted 15 June 2011.

1. Norman, M. R. The challenge of unconventional superconductivity. *Science* **332**, 196–200 (2011).
2. Armitage, N. P., Fournier, P. & Greene, R. L. Progress and perspectives on the electron-doped cuprates. *Rev. Mod. Phys.* **82**, 2421–2487 (2010).
3. Weber, C., Haule, K. & Kotliar, G. Strength of correlations in electron- and hole-doped cuprates. *Nature Phys.* **6**, 574–578 (2010).
4. Senechal, D. & Tremblay, A.-M. S. Hot spots and pseudogaps for hole- and electron-doped high-temperature superconductors. *Phys. Rev. Lett.* **92**, 126401 (2004).
5. Doiron-Leyraud, N. *et al.* Correlation between linear resistivity and  $T_c$  in the Bechgaard salts and the pnictide superconductor  $\text{Ba}(\text{Fe}_{1-x}\text{Co}_x)_2\text{As}_2$ . *Phys. Rev. B* **80**, 214531 (2009).
6. Löhneysen, H., v., Rosch, A., Vojta, M. & Wölfle, P. Fermi-liquid instabilities at magnetic quantum phase transitions. *Rev. Mod. Phys.* **79**, 1015–1075 (2007).
7. Moriya, T. & Ueda, K. Spin fluctuations and high temperature superconductivity. *Adv. Phys.* **49**, 555–606 (2000).
8. Sachdev, S. & Keimer, B. Quantum criticality. *Phys. Today* **64**, 29–35 (2011).
9. Rosch, A. Magnetotransport in nearly antiferromagnetic metals. *Phys. Rev. B* **62**, 4945–4962 (2000).
10. Bourbonnais, C. & Sedeki, A. Link between antiferromagnetism and superconductivity probed by nuclear spin relaxation in organic conductors. *Phys. Rev. B* **80**, 085105 (2009).
11. Taillefer, L. Scattering and pairing in cuprate superconductors. *Annu. Rev. Cond. Matter Phys.* **1**, 51–70 (2010).
12. Fournier, P. *et al.* Insulator-metal crossover near optimal doping in  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_4$ : Anomalous normal-state low temperature resistivity. *Phys. Rev. Lett.* **81**, 4720–4723 (1998).
13. Dagan, Y. *et al.* Evidence for a quantum phase transition in  $\text{Pr}_{2-x}\text{Ce}_x\text{CuO}_{4-\delta}$ . *Phys. Rev. Lett.* **92**, 167001 (2004).
14. Matsui, H. *et al.* Evolution of the pseudogap across the magnet-superconductor phase boundary of  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$ . *Phys. Rev. B* **75**, 224514 (2007).
15. Helm, T. *et al.* Evolution of the Fermi surface of the electron-doped high-temperature superconductor  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_4$  revealed by Shubnikov-de Haas oscillations. *Phys. Rev. Lett.* **103**, 157002 (2009).
16. Sawa, A. *et al.* Electron-doped superconductor  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$ : preparation of thin films and modified doping range for superconductivity. *Phys. Rev. B* **66**, 014531 (2002).
17. Jin, K. *et al.* Normal-state transport in electron-doped  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$  thin films in magnetic fields up to 40 Tesla. *Phys. Rev. B* **77**, 172503 (2008).
18. Jin, K. *et al.* Evidence for antiferromagnetic order in  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$  from angular magnetoresistance measurements. *Phys. Rev. B* **80**, 012501 (2009).
19. Jin, K. *et al.* Low-temperature Hall effect in electron-doped superconducting  $\text{La}_{2-x}\text{Ce}_x\text{CuO}_4$  thin films. *Phys. Rev. B* **78**, 174521 (2008).
20. Cooper, R. A. *et al.* Anomalous criticality in the electrical resistivity of  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Science* **323**, 603–607 (2009).
21. Daou, R. *et al.* Linear temperature dependence of resistivity and change in the Fermi surface at the pseudogap critical point of a high- $T_c$  superconductor. *Nature Phys.* **5**, 31–34 (2009).
22. Motoyama, E. M. *et al.* Spin correlations in the electron-doped high-transition-temperature superconductor  $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4\pm\delta}$ . *Nature* **445**, 186–189 (2007).
23. Lin, J. & Millis, A. J. Theory of low-temperature Hall effect in electron-doped cuprates. *Phys. Rev. B* **72**, 214506 (2005).
24. Fujita, M. *et al.* Low-energy spin fluctuations in the ground states of electron-doped  $\text{Pr}_{1-x}\text{LaCe}_x\text{CuO}_{4-\delta}$  cuprate superconductors. *Phys. Rev. Lett.* **101**, 107003 (2008).
25. Nakamae, S. *et al.* Electronic ground state of heavily overdoped nonsuperconducting  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. B* **68**, 100502 (2003).
26. Kubo, Y., Shimakawa, Y., Manako, T. & Igarashi, H. Transport and magnetic properties of  $\text{Ti}_2\text{Ba}_2\text{CuO}_{6+\delta}$  showing a  $\delta$ -dependent gradual transition from an 85-K superconductor to a nonsuperconducting metal. *Phys. Rev. B* **43**, 7875–7882 (1991).
27. Scalapino, D. J. The case for  $d_{x^2-y^2}$  pairing in the cuprate superconductors. *Phys. Rep.* **250**, 329–365 (1995).
28. Dhokarh, D. D. & Chubukov, A. V. Self-consistent Eliashberg theory,  $T_c$ , and the gap function in electron-doped cuprates. *Phys. Rev. B* **83**, 064518 (2011).
29. Monthoux, P., Pines, D. & Lonzarich, G. G. Superconductivity without phonons. *Nature* **450**, 1177–1183 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. Taillefer for extensive discussions and N. Doiron-Leyraud for some preliminary analysis of our zero-field data. We also appreciate discussions with A. Chubukov, A. Millis and C. Varma. Some experimental help was provided by X. Zhang, P. Bach and G. Droulers. This research was supported by the NSF under DMR-0952716 (J.P. and K.K.) and DMR-0653535 (R.L.G.) and the Maryland Center for Nanophysics and Advanced Materials (K.J. and N.P.B.).

**Author Contributions** K.J. prepared and characterized the thin-film samples. K.J., N.P.B., K.K. and J.P. performed the transport measurements and data analysis. N.P.B., J.P. and R.L.G. wrote the manuscript. R.L.G. conceived and directed the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.L.G. ([rgreene@squid.umd.edu](mailto:rgreene@squid.umd.edu)).

# Decoherence in crystals of quantum molecular magnets

S. Takahashi<sup>1,2,3</sup>, I. S. Tupitsyn<sup>4,5</sup>, J. van Tol<sup>6</sup>, C. C. Beedle<sup>7†</sup>, D. N. Hendrickson<sup>7</sup> & P. C. E. Stamp<sup>4,5</sup>

Quantum decoherence is a central concept in physics. Applications such as quantum information processing depend on understanding it; there are even fundamental theories proposed that go beyond quantum mechanics<sup>1–3</sup>, in which the breakdown of quantum theory would appear as an ‘intrinsic’ decoherence, mimicking the more familiar environmental decoherence processes<sup>4</sup>. Such applications cannot be optimized, and such theories cannot be tested, until we have a firm handle on ordinary environmental decoherence processes. Here we show that the theory for insulating electronic spin systems can make accurate and testable predictions for environmental decoherence in molecular-based quantum magnets<sup>5</sup>. Experiments on molecular magnets have successfully demonstrated quantum-coherent phenomena<sup>6–8</sup> but the decoherence processes that ultimately limit such behaviour were not well constrained. For molecular magnets, theory predicts three principal contributions to environmental decoherence: from phonons, from nuclear spins and from intermolecular dipolar interactions. We use high magnetic fields on single crystals of Fe<sub>8</sub> molecular magnets (in which the Fe ions are surrounded by organic ligands) to suppress dipolar and nuclear-spin decoherence. In these high-field experiments, we find that the decoherence time varies strongly as a function of temperature and magnetic field. The theoretical predictions are fully verified experimentally, and there are no other visible decoherence sources. In these high fields, we obtain a maximum decoherence quality-factor of  $1.49 \times 10^6$ ; our investigation suggests that the environmental decoherence time can be extended up to about 500 microseconds, with a decoherence quality factor of  $\sim 6 \times 10^7$ , by optimizing the temperature, magnetic field and nuclear isotopic concentrations.

Environmental decoherence processes are reasonably well understood at the atomic scale<sup>9</sup> (although some poorly understood noisy sources remain<sup>10</sup>). However both quantum information processing, and the fundamental tests noted above, require an understanding of decoherence in larger systems, where experimental decoherence rates are usually much larger than theoretical predictions. This discrepancy is usually attributed to ‘extrinsic’ sources (external noise, uncontrolled disorder/impurities). We thus need to find systems, with many degrees of freedom, where extrinsic decoherence can be eliminated, and where we have a quantitative understanding of other decoherence sources.

Many insulating electronic spin systems are currently the subject of intense experimental interest, notably in semiconductor quantum dots<sup>11,12</sup>, nitrogen-vacancy centres in diamond<sup>13–15</sup> and large-spin magnetic molecules<sup>6–8</sup>. In all these systems, three environmental decoherence mechanisms are involved. The electronic spins couple locally to (1) phonons (an oscillator bath<sup>16</sup>); (2) to large numbers of nuclear spins (a spin bath<sup>17</sup>); and (3) to each other via dipolar interactions. The long range of dipolar interactions is a major problem: it makes quantum error correction more difficult, is theoretically complicated<sup>18</sup> and is very hard to eliminate experimentally.

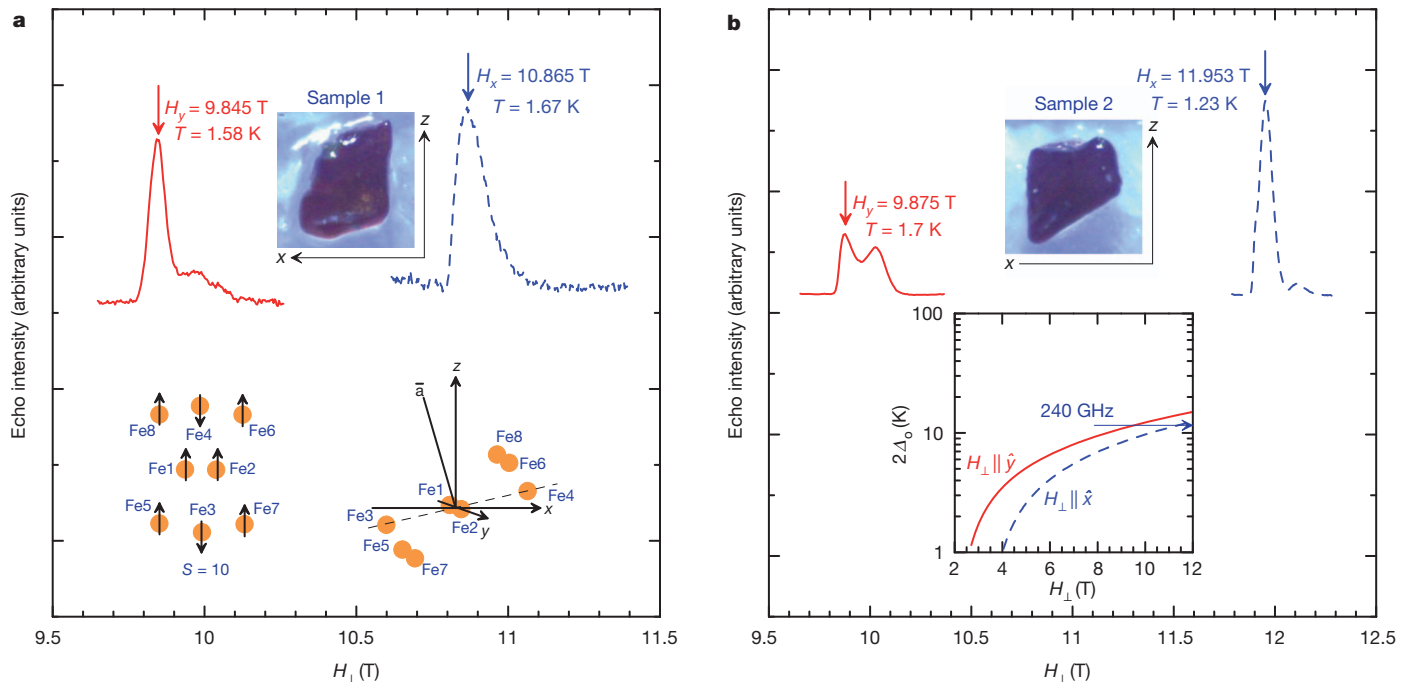
In these crystalline Fe<sub>8</sub> molecular magnets, the electronic spins are structurally ordered, and quantum coherence is observed in the collective (magnon) motion of the spins, rather than in single qubit dynamics. Two great advantages of the Fe<sub>8</sub> system<sup>19</sup> are that the interaction strengths are well known, allowing quantitative predictions, and that it can be prepared with little disorder and few impurities, reducing the danger of extrinsic decoherence. The number of relevant environmental degrees of freedom is very large; depending on isotopic concentrations, there are  $10^{50}$ – $10^{54}$  nuclear spin levels in each molecule, and the system couples to a bulk phonon bath.

In the spin echo experiments described here, a Hahn echo sequence<sup>20</sup> was created in two single crystals of Fe<sub>8</sub> molecules, with natural isotopic concentrations, using a 240-GHz pulsed ESR (electron spin resonance) spectrometer<sup>21,22</sup>. Thus a uniform ESR precession mode (a  $\mathbf{k} = 0$  magnon, where  $\mathbf{k}$  is the wave vector of the magnon) interacts with its surroundings, and we measure the decoherence time  $T_2$ . At low temperature, each electronic spin system behaves as a two-level quantum bit (qubit), with a splitting  $2\Delta_0$  that depends strongly on the local transverse field  $\mathbf{H}_\perp$ , perpendicular to the easy axis  $\hat{z}$  (see Fig. 1 inset). Almost all previous experiments on electron spin systems examined the low-field regime, where nuclear spin decoherence is very strong; here we go to the high-field regime, where its effects are much weaker. Typical ESR results are shown in Fig. 1; we discuss them in detail below.

To understand  $T_2$ , and the ESR lineshape, we need to look at the processes contributing to them. For convenience, we define a dimensionless decoherence rate  $\gamma_\phi = \hbar/\Delta_0 T_2$  (where  $\hbar$  is Planck’s constant  $h$  divided by  $2\pi$ ) and the associated ‘decoherence Q-factor’,  $Q_\phi = \pi/\gamma_\phi$ . Then the processes contributing to  $\gamma_\phi$  are as follows (the full quantitative discussion, for the two samples in this experiment, is given in Supplementary Information):

First, nuclear spins interact locally with each molecular spin, and cause decoherence by a ‘motional narrowing’ process in which they attempt to entangle with the fast-moving qubit<sup>18,23</sup>. The nuclear decoherence rate is  $\gamma_\phi^N = E_0^2/2\Delta_0^2$ , where  $E_0$  is the half-width of the Gaussian multiplet of nuclear spin states coupled to the qubit; and the nuclear contribution to the ESR linewidth is just  $E_0$ . Now in this experiment, with naturally occurring isotopic concentrations,  $E_0 = 4 \times 10^{-4}$  K at these fields, where  $\Delta_0 = 5.75$  K. Thus  $\gamma_\phi^N \approx 10^{-9}$  is very small, simply because  $\Delta_0$  is so large in these fields; and the nuclear spin contribution ( $\sim E_0$ ) to the linewidth is also very small compared to the main contributions. Isotopic substitution of deuterium for the 120 protons in each molecule will further decrease  $\gamma_\phi^N$  by a factor of 15.2 to  $\gamma_\phi^N \approx 7 \times 10^{-11}$ . In principle, there can be a ‘noise’ contribution from the intrinsic nuclear dynamics caused by internuclear interactions<sup>17</sup>; however, in contrast to quantum dot systems<sup>24</sup>, such contributions are very small in molecular magnets<sup>23</sup> (even in systems with strongly interacting nuclei like Mn<sub>12</sub>, where they have been seen<sup>25</sup>).

<sup>1</sup>Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA. <sup>2</sup>Institute for Terahertz Science and Technology, University of California, Santa Barbara, California 93106, USA. <sup>3</sup>Department of Physics, University of California, Santa Barbara, California 93106, USA. <sup>4</sup>Pacific Institute of Theoretical Physics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>5</sup>Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>6</sup>National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA. <sup>7</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA. <sup>†</sup>Present address: National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA.



**Figure 1** | Typical ESR spectra, showing echo intensity as a function of transverse magnetic field,  $H_{\perp}$ . Data are shown for two different samples, at different temperatures and orientations of field, and at  $\omega_{\text{ESR}} = 240$  GHz. **a**, Sample 1. Solid red line,  $H_{\perp} \parallel \hat{y}$ ,  $T = 1.58$  K; dashed blue line,  $H_{\perp} \parallel \hat{x}$ ,  $T = 1.67$  K. Top inset, sample dimensions are approximately  $z: x: y = 1,000: 700: 250$   $\mu\text{m}$ . Lower left inset, the low- $T$  spin structure of the  $\text{Fe}_8$  molecule.

Second, the form of the local spin–phonon interaction is determined by the system symmetry. At high fields this interaction simplifies, and we find a dimensionless phonon decoherence rate<sup>18</sup> given by  $\gamma_{\phi}^{\text{ph}} = [(\mathcal{F}_{\text{AS}} \Delta_o^2) / (\pi \rho c_s^3 \hbar^3)] \coth(\Delta_o / k_B T)$ , where  $\rho$  is the sample density,  $c_s$  the sound velocity, and  $\mathcal{F}_{\text{AS}}$  the relevant spin–phonon matrix element. The contribution of this spin–phonon process to the ESR linewidth is negligible.

Third, the intermolecular dipole interaction directly couples the  $\mathbf{k} = 0$  ESR precession mode to finite-momentum magnons; it may decay spontaneously into multiple magnons, or scatter off existing thermal magnons. This process affects the ESR lineshape and the decoherence rate very differently. The long-range dipolar interaction creates a distribution of demagnetization fields around the sample. In highly polarized samples, this is strongly sample-shape dependent, but for annealed samples, it is Gaussian distributed<sup>26,27</sup>; in both cases it can be calculated numerically. The lineshape then reflects the quite broad distribution of these fields. However, decoherence comes from the magnon decay process described above, and depends only on the phase space available for these processes at the resonance field; it can then be calculated directly from the analytic expression for this process. At the experimental temperatures, the magnon decoherence rate is  $\sim \exp(-2\Delta_o / k_B T)$ , where  $k_B$  is Boltzmann’s constant and  $T$  is temperature, coming almost entirely from thermal magnon scattering.

Last, there can be extrinsic contributions from impurities and defects, which typically cause the easy-axis anisotropy parameter  $D$  of the  $\text{Fe}_8$  Hamiltonian to fluctuate around the sample (the ‘ $D$ -strain’ effect<sup>28,29</sup>). This will then contribute to the ESR linewidth. Such static impurities and defects can cause  $\Delta_o$  to vary in the sample (although we find no evidence for such a spread in this experiment). However they can not contribute to decoherence at all, provided they are static, because then they simply shift the individual qubit energies. On the other hand any impurities or defects with significant dynamics will cause extrinsic decoherence.

In Fig. 2 we show how these contributions to the decoherence rate are predicted to vary with field and temperature for a crystalline  $\text{Fe}_8$

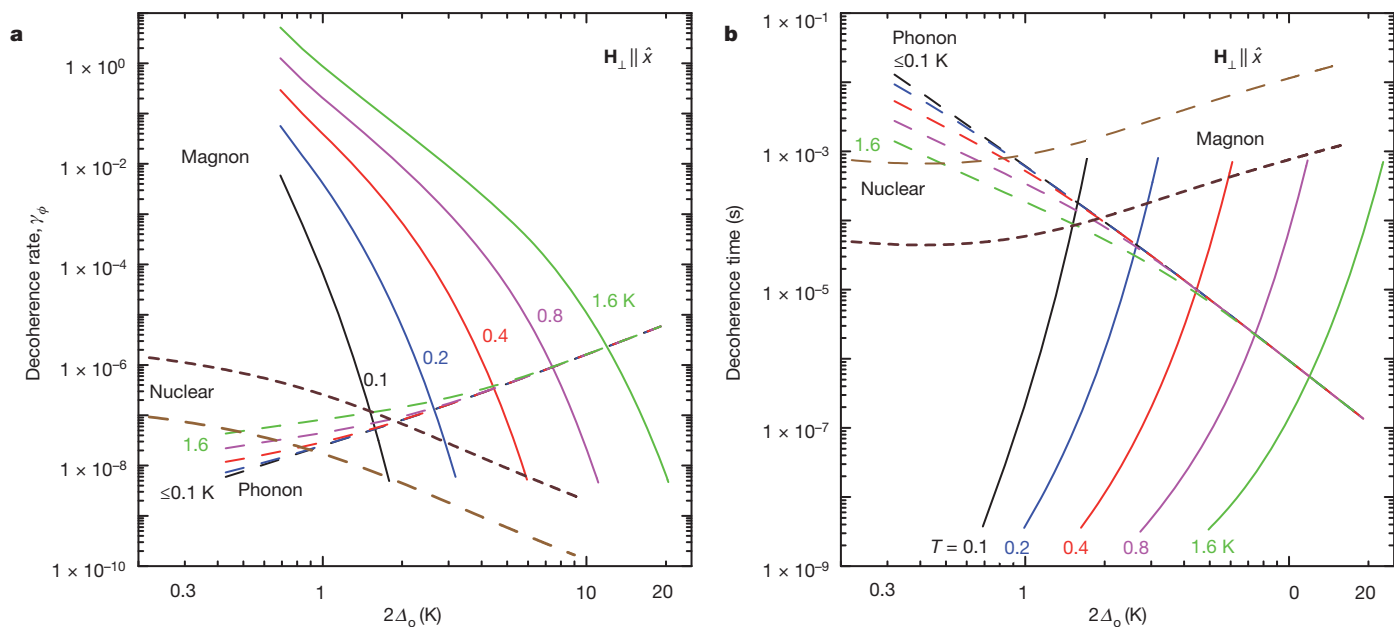
system. The spin–phonon contribution  $\gamma_{\phi}^{\text{ph}}$  increases with applied transverse field, because the available phonon phase space increases sharply with  $\Delta_o$ . However the nuclear spin decoherence rate decreases with field: roughly  $\gamma_{\phi}^{\text{N}} \propto 1/\Delta_o^2$ , because  $E_o$  changes quite slowly with field (and also decreases as  $\Delta_o$  increases). Thus there is a crossover, with a minimum  $\gamma_{\phi} \approx 10^{-7}$  when  $\Delta_o \approx 1$  K. However these two ‘single-qubit’ decoherence mechanisms are entirely masked for a dense crystal by the dipolar ‘magnon’ decoherence, except at high fields (where dipolar decoherence competes with phonon decoherence) or at very low temperatures and low fields (where it competes with nuclear spin decoherence). In this experiment, we chose to go to high fields.

With all this in mind, we return to the ESR results obtained by echo-detected field sweep, in Fig. 1. The resonant peaks are broadened, with a width  $\sim 0.1$  T; the peculiar structure of the peak when  $H_{\perp} \parallel \hat{y}$ , discussed in detail in Supplementary Information, comes from dipolar interactions. These ESR signals may be understood as follows. The qubit splitting  $2\Delta_o$  varies with field as shown in Fig. 1b inset. For fields  $H_y = 9.5$  T,  $H_x = 0$ , or  $H_x = 11.3$  T,  $H_y = 0$ , the electronic spin Hamiltonian<sup>18,30</sup> for  $\text{Fe}_8$  predicts  $2\Delta_o(H_{\perp}) \approx 11.5$  K, equivalent to our spectrometer frequency of 240 GHz (see Fig. 1 inset), implying we should see resonance peaks at these fields. These predictions are reasonably well satisfied in both samples. The discrepancies, discussed in detail in Supplementary Information, come from two sources: (1) sample misorientation, and (2) weak departures at high field from the model Hamiltonian<sup>18,30</sup> used to predict the field splitting.

The results of the measurements for each sample, together with the calculated theoretical decoherence times for  $\text{Fe}_8$ , are presented in Fig. 3. The agreement is very good; we emphasize that apart from the size of the spin–phonon coupling, which is not known exactly, there are no adjustable parameters in these fits. The decoherence times and rates in the experiment range over roughly an order of magnitude, with a maximum  $T_2 \approx 0.63$   $\mu\text{s}$ , corresponding to  $Q_{\phi} \approx 1.49 \times 10^6$ , at the lowest temperature we went to.

A number of features should be stressed here. First, notice how differently the decoherence and the ESR lineshape are affected by





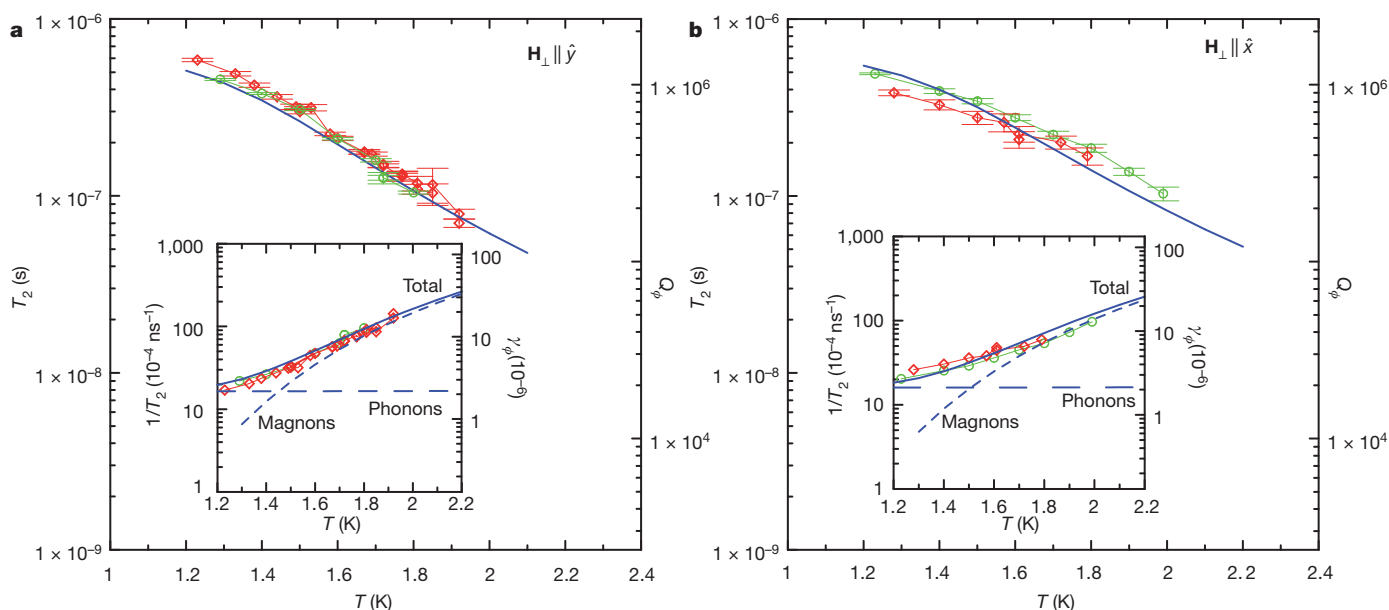
**Figure 2 | Calculated contributions to the decoherence coming from the coupling to nuclear spins, phonons and magnons.** **a**, The three individual contributions which sum to give the dimensionless decoherence rate  $\gamma_\phi = \hbar/T_2\Delta_0$ , as a function of the qubit splitting in the case  $H_\perp \parallel \hat{x}$ . **b**, The three corresponding contributions to the decoherence time,  $T_2$ . In both panels:

the different environmental couplings. The ESR linewidth and line-shape are completely dominated by static impurity fields and by the spatially varying dipolar fields. However the decoherence is completely dominated, at these high fields, by the phonons and dipolar interactions. At lower fields, the nuclear spin decoherence would also be important, but its effect on the ESR linewidth would still be negligible. Note also that whereas the dipolar contribution to the ESR lineshape depends strongly on sample shape, this shape can only affect the

brown dashed lines, nuclear contribution (short-dashed brown lines are for the natural isotopic concentrations, long-dashed brown lines for the deuterated system); solid lines of different colours, magnon contributions at different temperatures (shown) from 0.1 to 1.6 K; long-dashed lines of different colours, phonon contributions, shown for the same temperatures as in the magnon case.

dipolar decoherence near the edges of this line. In the middle of the line, when the decoherence is coming from molecules near the centre of the sample where the field is homogeneous, one expects no dependence of the dipolar decoherence on sample shape. This is also what we found in the experiment.

Second, we emphasize how the experiment tests the phonon and dipolar contributions to the decoherence separately: they have very different temperature dependences in the regime covered here, with



**Figure 3 | Measured and calculated decoherence times  $T_2$  in samples 1 and 2, as a function of temperature.** **a**, Results for  $H_\perp \parallel \hat{y}$ . Main panel: thin red line with diamonds, measured using sample 1,  $H_y = 9.845$  T; thin green line with circles, measured using sample 2,  $H_y = 9.875$  T; vertical and horizontal error bars, standard errors of  $T_2$  data fits and uncertainty in temperature ( $\Delta T = \pm 0.05$  K), respectively. Thick blue line, calculations including phonon and magnon contributions,  $H_y = 9.5$  T. Inset: partial contributions calculated

for  $T_2^{-1}$  (solid line) from magnons (dashed line) and phonons (long-dashed line), together with the corresponding experimental results for the two samples (diamonds and circles). The scale on the right-hand side of the main panel indicates the decoherence  $Q$ -factor,  $Q_\phi = \pi/\gamma_\phi = \pi T_2\Delta_0/\hbar$ ; the right-hand scale on the inset shows  $\gamma_\phi$ . **b**, As for **a**, but now for  $H_\perp \parallel \hat{x}$ . The experimental curves were measured at  $H_x = 10.865$  T (sample 1) and  $H_x = 11.953$  T (sample 2). The theoretical curves are obtained at  $H_x = 11.3$  T.

phonons dominating below  $\sim 1.2$  K, and magnons dominating at higher temperature. We find agreement between theory and experiment, with no adjustable parameters, across roughly an order of magnitude in decoherence rate. Thus all decoherence in the experiment can be accounted for by environmental sources. This implies that we have no measurable extrinsic decoherence here, either from disorder or dynamic impurities. Nor do we have evidence for any other contributions, either from “third-party decoherence”<sup>4</sup>, or from any of the ‘intrinsic decoherence’ sources<sup>1</sup> discussed in the literature.

Third, we note a key difference between the coherence here, which involves a macroscopic number of qubits excited coherently in a spin wave, and that in other qubit systems, in which entanglement is achieved locally, involving just one or a few qubits. The reason we can do this is because we are dealing with a single crystal.

Last, the present investigation suggests that one can optimize the qubit decoherence  $T_2$  and  $Q$ -factor  $Q_\phi$ , as a function of field and temperature, using the results plotted in Fig. 2. We see that lower temperature allows use of a smaller ESR frequency  $2A_0$ ; these two changes strongly reduce the dipolar and phonon decoherence contributions, giving a large increase in  $T_2$ , and a somewhat smaller increase in  $Q_\phi$ . The optimal decoherence rate comes when the phonon and nuclear spin decoherence contributions cross (for natural isotopic concentrations, this is at  $2A_0 \approx 2$  K), provided also that  $T < 0.13$  K, so that the dipolar/magnon decoherence can be ignored. One then finds that  $\gamma_\phi \approx 1.5 \times 10^{-7}$ , so that  $Q_\phi = 2 \times 10^7$ , corresponding to a decoherence time  $T_2 \approx 50 \mu\text{s}$ . However with isotopic substitution of deuterons in place of the protons, the optimal decoherence time rises to  $T_2 \approx 500 \mu\text{s}$ , at  $2A_0 = 0.8$  K  $\equiv 17$  GHz, and  $T = 45$  mK. This corresponds to  $\gamma_\phi \approx 5 \times 10^{-8}$  and  $Q_\phi \approx 6 \times 10^7$ . These considerations show the usefulness of this kind of theory in the optimal design of spin qubit systems—notice the crucial importance of controlling the dipolar interactions between qubits. Notice also that if quantum mechanics is to be tested on anything but microscopic scales, it will be essential to continue developing theory and experiment for systems like the present one, where the environmental decoherence processes can be understood quantitatively, and where extrinsic decoherence sources can be largely eliminated.

## METHODS SUMMARY

The single molecule characteristics were calculated using previous results for crystal field parameters<sup>30</sup> and for the effect of high field on these<sup>18</sup>. Hyperfine couplings were taken from previous work<sup>18</sup>; the spin–phonon couplings were estimated using standard magnetostriction theory, in the high field regime. The dipolar fields were calculated numerically, taking into account the unit cell crystal structure, the sample shape and the field direction, for each sample. The nuclear spin and phonon decoherence rates were determined analytically, using standard methods<sup>16,17</sup>, using the previously determined hyperfine and spin–phonon couplings. The dipolar decoherence rate, from four-magnon scattering and decay processes, was determined numerically, for the given sample shape, field and crystal lattice structure, using analytic formulas<sup>18</sup> for the magnon spectrum and dipolar coupling functions.

Single crystals of  $\text{Fe}_8$  magnetic molecules were synthesized using the method of ref. 19. Each crystal was indexed and unit-cell parameters were checked to ensure consistency. Continuous-wave/pulsed ESR measurements were carried out using the 240-GHz ESR spectrometer at the National High Magnetic Field Laboratory (NHMFL)<sup>21,22</sup>. The system consists of a 12.5-T superconducting magnet, a 40-mW 240-GHz source, quasi-optics, a superheterodyne detection system, and a  $^4\text{He}$  flow cryostat. The spin decoherence time was measured by a Hahn echo sequence ( $\pi/2$ – $\tau$ – $\pi$ – $\tau$ -echo) where the delay  $\tau$  is varied<sup>20</sup>. The magnetic component of the 240-GHz pulses was perpendicular to the d.c. magnetic field, to generate coherent magnons in the sample; their duration was adjusted to maximize the echo signals, and was typically 200–300 ns.

Received 3 April; accepted 16 June 2011.

Published online 20 July 2011.

1. 't Hooft, G. Quantum gravity as a dissipative deterministic system. *Class. Quantum Gravity* **16**, 3263–3279 (1999).

2. Penrose, R. On gravity's role in quantum state reduction. *Gen. Relativ. Gravit.* **28**, 581–600 (1996).
3. Leggett, A. J. Testing the limits of quantum mechanics: motivation, state of play, prospects. *J. Phys. Condens. Matter* **14**, R415–R451 (2002).
4. Stamp, P. C. E. The decoherence puzzle. *Stud. Hist. Phil. Mod. Phys.* **37**, 467–497 (2006).
5. Gatteschi, D., Sessoli, R. & Villain, J. *Molecular Nanomagnets* (Oxford Univ. Press, 2006).
6. Takahashi, S. *et al.* Coherent manipulation and decoherence of  $S = 10$  single-molecule magnets. *Phys. Rev. Lett.* **102**, 087603 (2009).
7. Bertina, S. *et al.* Quantum oscillations in a molecular magnet. *Nature* **453**, 203–206 (2008).
8. Schlegel, C. *et al.* Direct observation of quantum coherence in single-molecule magnets. *Phys. Rev. Lett.* **101**, 147203 (2008).
9. Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
10. Labaziewicz, J. *et al.* Temperature dependence of electric field noise above gold surfaces. *Phys. Rev. Lett.* **101**, 180602 (2008).
11. Xu, X. *et al.* Optically controlled locking of the nuclear field via coherent dark-state spectroscopy. *Nature* **459**, 1105–1109 (2009).
12. Ladd, T. D. *et al.* Pulsed nuclear pumping and spin diffusion in a single charged quantum dot. *Phys. Rev. Lett.* **105**, 107401 (2010).
13. Takahashi, S. *et al.* Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.* **101**, 047601 (2008).
14. Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
15. Hanson, R. *et al.* Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
16. Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59**, 1–85 (1987).
17. Prokof'ev, N. V. & Stamp, P. C. E. Theory of the spin bath. *Rep. Prog. Phys.* **63**, 669–726 (2000).
18. Morello, A., Stamp, P. C. E. & Tupitsyn, I. S. Pairwise decoherence in coupled spin qubit networks. *Phys. Rev. Lett.* **97**, 207206 (2006).
19. Wiegand, K., Pohl, K., Jibril, I. & Huttner, G. Hydrolysis products of the monomeric amine complex  $(\text{C}_6\text{H}_{15}\text{N}_3)\text{FeCl}_3$ : the structure of the octameric iron(III) cation of  $[(\text{C}_6\text{H}_{15}\text{N}_3)_6\text{Fe}_8(\mu_3\text{-O})_2(\mu_2\text{-OH})_{12}]\text{Br}_7 \cdot 8\text{H}_2\text{O}$ . *Angew. Chem. Int. Edn Engl.* **23**, 77–78 (1984).
20. Hahn, E. L. Spin echoes. *Phys. Rev.* **80**, 580–594 (1950).
21. van Tol, J., Brunel, L. C. & Wylde, R. J. A quasi-optical transient electron spin resonance spectrometer operating at 120 and 240 GHz. *Rev. Sci. Instrum.* **76**, 074101 (2005).
22. Morley, G. W., Brunel, L. C. & van Tol, J. A. Multifrequency high-field pulsed electron paramagnetic resonance/electron-nuclear double resonance spectrometer. *Rev. Sci. Instrum.* **79**, 064703 (2008).
23. Stamp, P. C. E. & Tupitsyn, I. S. Coherence window in the dynamics of quantum nanomagnets. *Phys. Rev. B* **69**, 014401 (2004).
24. Yao, W., Liu, R. B. & Sham, L. J. Theory of electron spin decoherence by interacting spins in a quantum dot. *Phys. Rev. B* **74**, 195301 (2006).
25. Morello, A., Bakharev, O. N., Brom, H. B., Sessoli, R. & de Jongh, L. J. Nuclear Spin Dynamics in the Quantum regime of a single molecule magnet. *Phys. Rev. Lett.* **93**, 197202 (2004).
26. Cuccoli, A. *et al.* Dipolar interaction and incoherent quantum tunneling: a Monte Carlo study of magnetic relaxation. *Europhys. J. B* **12**, 39–46 (1999).
27. Tupitsyn, I. S., Stamp, P. C. E. & Prokof'ev, N. V. Hole digging in ensembles of tunneling molecular magnets. *Phys. Rev. B* **69**, 132406 (2004).
28. Hill, S. *et al.* D-strain, g-strain, and dipolar interactions in the  $\text{Fe}_8$  and  $\text{Mn}_{12}$  single molecule magnets: an EPR lineshape analysis. *Int. J. Mod. Phys. B* **16**, 3326–3329 (2002).
29. Park, K. *et al.* Effects of D-strain, g-strain, and dipolar interactions on EPR linewidths of the molecular magnets  $\text{Fe}_8$  and  $\text{Mn}_{12}$ . *Phys. Rev. B* **65**, 014426 (2001).
30. Barra, A. L., Gatteschi, D. & Sessoli, R. High-frequency EPR spectra of  $[\text{Fe}_8\text{O}_2(\text{H}_2\text{O})_{12}(\text{tacn})_6]\text{Br}_8$ : a critical appraisal of the barrier for the reorientation of the magnetization in single-molecule magnets. *Chem. Eur. J.* **6**, 1608–1614 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the NSF (DMR-0520481, DMR-0703925), the Keck Foundation (S.T. and J.v.T.), NSERC, CIFAR, PITP, the John E. Fetzer Memorial Trust (grant D21-C62) and the Center for Philosophy and the Natural Sciences, California State University, Sacramento (I.S.T. and P.C.E.S.). The National High Magnetic Field Laboratory is supported by NSF Cooperative Agreement DMR-0654118, by the State of Florida, and by the DOE.

**Author Contributions** S.T., I.S.T. and P.C.E.S. contributed to the writing of the manuscript. S.T., I.S.T. and P.C.E.S. conceived the ESR experiments. The ESR measurements were carried out by S.T. and J.v.T. The theoretical work was done by I.S.T. and P.C.E.S. C.C.B. and D.N.H. synthesized  $\text{Fe}_8$  crystals and performed X-ray diffraction measurements.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.T. ([susumuta@usc.edu](mailto:susumuta@usc.edu)) or P.C.E.S. ([stamp@phas.ubc.ca](mailto:stamp@phas.ubc.ca)).

# Sensitivity of coccolithophores to carbonate chemistry and ocean acidification

L. Beaufort<sup>1</sup>, I. Probert<sup>2</sup>, T. de Garidel-Thoron<sup>1</sup>, E. M. Bendif<sup>2</sup>, D. Ruiz-Pino<sup>3</sup>, N. Metz<sup>1</sup>, C. Goyet<sup>4</sup>, N. Buchet<sup>1</sup>, P. Coupel<sup>3</sup>, M. Grelaud<sup>1†</sup>, B. Rost<sup>5</sup>, R. E. M. Rickaby<sup>6</sup> & C. de Vargas<sup>2</sup>

About one-third of the carbon dioxide (CO<sub>2</sub>) released into the atmosphere as a result of human activity has been absorbed by the oceans<sup>1</sup>, where it partitions into the constituent ions of carbonic acid. This leads to ocean acidification, one of the major threats to marine ecosystems<sup>2</sup> and particularly to calcifying organisms such as corals<sup>3,4</sup>, foraminifera<sup>5–7</sup> and coccolithophores<sup>8</sup>. Coccolithophores are abundant phytoplankton that are responsible for a large part of modern oceanic carbonate production. Culture experiments investigating the physiological response of coccolithophore calcification to increased CO<sub>2</sub> have yielded contradictory results between and even within species<sup>8–11</sup>. Here we quantified the calcite mass of dominant coccolithophores in the present ocean and over the past forty thousand years, and found a marked pattern of decreasing calcification with increasing partial pressure of CO<sub>2</sub> and concomitant decreasing concentrations of CO<sub>3</sub><sup>2–</sup>. Our analyses revealed that differentially calcified species and morphotypes are distributed in the ocean according to carbonate chemistry. A substantial impact on the marine carbon cycle might be expected upon extrapolation of this correlation to predicted ocean acidification in the future. However, our discovery of a heavily calcified *Emiliania huxleyi* morphotype in modern waters with low pH highlights the complexity of assemblage-level responses to environmental forcing factors.

To assess the influence of the environment on coccolithophore calcification, we investigated 180 surface-water and 555 sediment-core samples encompassing a wide spectrum of present and past oceanic conditions (Fig. 1). The family Noëlaerhabdaceae (including the extant genera *Emiliania*, *Gephyrocapsa* and *Reticulofenestra*) has dominated coccolithophore communities numerically for more than 20 million years. Although they are closely related genetically<sup>12</sup>, noëlaerhabdaceans vary both between and within species in the morphology of the calcite scales (coccoliths) that form their composite skeletons (coccospheres)<sup>13</sup>. An optical method for automatic analysis of the size and mass of individual noëlaerhabdacean coccospheres and detached coccoliths was applied, with an average of 700 coccoliths measured per sample. The mass of isolated coccoliths was strongly correlated with the mass of coccospheres in modern samples ( $R^2 = 0.88$ ), demonstrating the validity of coccolith mass as a proxy for the calcification state of noëlaerhabdaceans.

Temperature, salinity, alkalinity and dissolved inorganic carbon were recorded directly from modern water samples, allowing the derivation of all carbonate-chemistry parameters<sup>14</sup>. To reconstruct the carbonate chemistry of surface waters in the past, published palaeo-proxy records for sea surface temperature and salinity from each core site were combined with records of CO<sub>2</sub> partial pressure ( $p_{\text{CO}_2}$ ) from Antarctic ice. Uncertainties due to the propagation of errors from temperature and salinity estimates, and from our assumption of stability over the past 40,000 years (kyr) in relationships linking

alkalinity and salinity, the ratio of stable oxygen isotopes ( $\delta^{18}\text{O}$ ) in water and salinity, and  $p_{\text{CO}_2}$  in water and in the atmosphere, had only a limited effect on the general trend (Supplementary Information). The resulting glacial–interglacial ranges of carbonate-chemistry parameters are similar to those published elsewhere<sup>5,15</sup>.

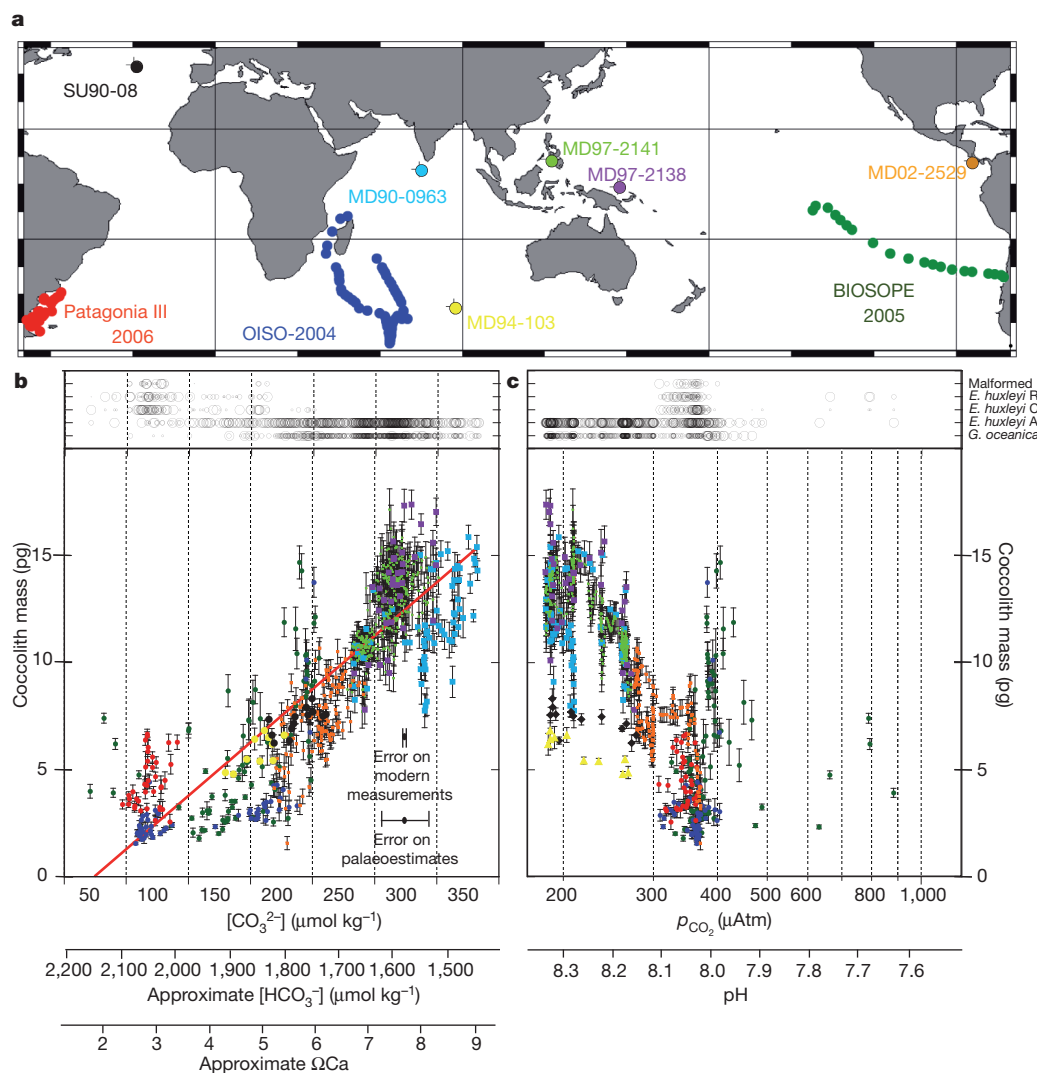
Temperature, salinity, light and nutrients have all been reported to affect coccolithophore calcification or coccolith size<sup>13,16–20</sup>. Analysis of our large data set reveals that temperature and salinity are not strongly correlated with coccolith mass. The contrasting correlations between coccolith mass and temperature in modern samples ( $R = 0.59$ ) and in sediment samples ( $R$  between  $-0.12$  and  $-0.71$ ) reflect the differing relationship between temperature and carbonate concentration ( $[\text{CO}_3^{2-}]$ ) in modern and past oceans (Table 1). No significant correlations were found between coccolith mass and productivity-related parameters (chlorophyll or cell abundance), where these were available.

Coccolith mass was related to carbonate chemistry. Significant overall correlations of coccolith mass with pH and  $p_{\text{CO}_2}$  were recorded, but with notable regional variations (Fig. 1c), indicating that these parameters are not solely responsible for the observed trend. The only correlations that were highly significant in all subsets of the data were those linking coccolith mass to  $[\text{CO}_3^{2-}]$ ,  $[\text{HCO}_3^-]$  and calcite saturation state ( $\Omega_{\text{Ca}}$ ) ( $R^2 > 0.74$ , Fig. 1b and Table 1). The influence of carbonate chemistry was particularly notable in sediment records: during the Last Glacial Maximum,  $p_{\text{CO}_2}$  was low ( $[\text{CO}_3^{2-}]$  was high) and coccolith mass was high (Fig. 2). During deglaciation and the Holocene epoch, coccolith mass decreased with increasing  $p_{\text{CO}_2}$  at all latitudes in different ocean basins. This trend cannot have resulted from post-depositional corrosion because glacial–interglacial dissolution conditions evolved in opposite directions in the Atlantic and Pacific oceans<sup>21</sup>. A significant correlation was also found between  $[\text{CO}_3^{2-}]$  and amount of calcite per unit cell volume ( $R^2 = 0.76$ ), indicating that the degree of calcification was not dependent on cell size. There was a negative correlation between calcite mass and  $[\text{HCO}_3^-]$ , despite this anion being the primary carbon source for intracellular calcification in *E. huxleyi*<sup>16,22,23</sup>.  $[\text{HCO}_3^-]$  varied by about 20% in our data set, but it is the most abundant carbon species in sea water and is therefore unlikely to limit biomineralization. In contrast,  $[\text{CO}_3^{2-}]$  varied by 77%. Several physico-chemical parameters could synergistically affect calcification, but multiple regression did not markedly increase the significance of the correlation of coccolith mass with  $[\text{CO}_3^{2-}]$ . Although various parameters clearly exert a localized influence on coccolith mass, our data point to  $[\text{CO}_3^{2-}]$  and  $\Omega_{\text{Ca}}$  being key parameters in the global assemblage-scale response of noëlaerhabdacean calcification to ocean acidification.

We next explored what drives the trend linking calcification and carbonate chemistry. The results of monoclonal culture experiments, showing a decrease in noëlaerhabdacean calcification (degree and rate) with ocean acidification (for example, refs 8 and 11) have focused

<sup>1</sup>CEREGE, CNRS/Université Aix-Marseille, Avenue L. Philibert BP80, 13545 Aix-en-Provence, Cedex 4, France. <sup>2</sup>Station Biologique de Roscoff, CNRS/Université P. & M. Curie, Place G. Teissier, 29680 Roscoff, France. <sup>3</sup>LOCEAN-IPSL, CNRS/Université P. & M. Curie, BP 100, 4 Place Jussieu, 75252 Paris, Cedex 5, France. <sup>4</sup>Université de Perpignan, 52 Avenue P. Alduy, 66860 Perpignan, Cedex 9, France. <sup>5</sup>Alfred Wegener Institute, Am Handelshafen 12, 27570 Bremerhaven, Germany. <sup>6</sup>Oxford University, Department of Earth Sciences, Parks Road, Oxford OX1 3OR, UK. <sup>†</sup>Present address: ICTA, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.





**Figure 1 | Relationships between coccolith mass and carbonate chemistry.** **a**, Map of sample locations. **b**, Relationship between  $[\text{CO}_3^{2-}]$  and coccolith mass (colours correspond to those on the map in **a**). Vertical bars are standard error on the mass distribution. For comparison, the corresponding scales for  $\Omega\text{Ca}$  and  $[\text{HCO}_3^-]$  are shown. Error bars for  $[\text{CO}_3^{2-}]$  are estimated from modern and past data. **c**, Relationship between coccolith mass and  $p\text{CO}_2$  in water. The corresponding pH scale is given. Upper panels in **b** and **c** show qualitative distribution of noëlaerhabdacean taxa: large circles for dominant taxa (>50%), medium circles for abundant taxa (30–50%) and small circles for less abundant taxa (10–30%).

attention on the existence of a direct, environmentally mediated physiological constraint on calcification. Coccoliths are secreted intracellularly in coccolith-producing vesicles in which pH and  $\Omega\text{Ca}$  are maintained at levels that stimulate calcite precipitation. Several explanations have been proposed to explain decreased calcification rates with ocean acidification<sup>24</sup>, for example, that the necessary outward transport of protons may become more costly in low-pH and/or low- $[\text{CO}_3^{2-}]$  waters<sup>25</sup>. The environmental relevance of such a physiological mechanism cannot be resolved directly from our data, but in culture experiments, the maximum decrease in mass of a clone over carbonate-chemistry ranges comparable to those of our data set is about 20%; only a fraction of the total response in our data set. In addition, some cultured strains of noëlaerhabdaceans and other coccolithophores are capable of maintaining calcification (degree and/or

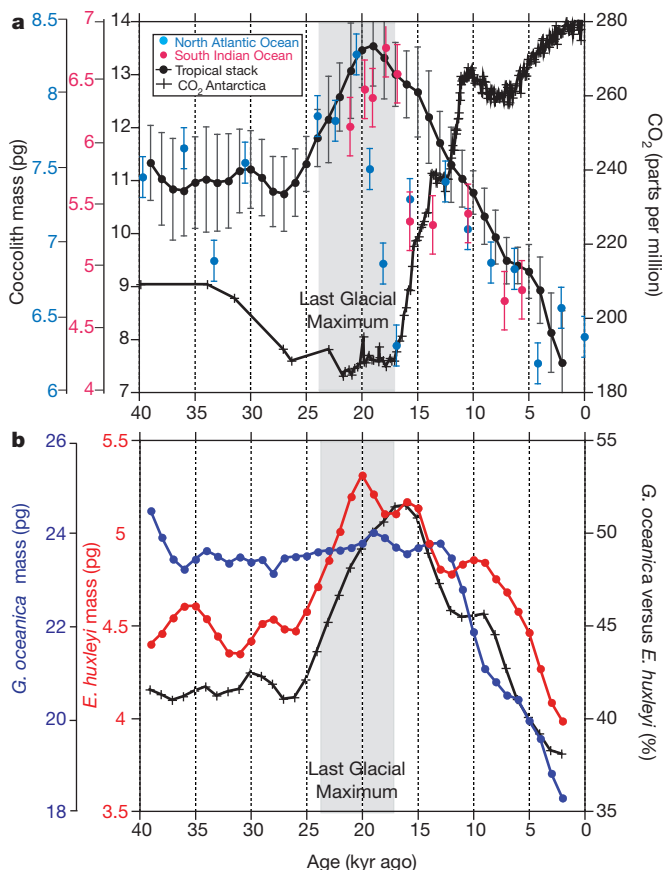
rate) over certain carbonate-chemistry ranges, a phenomenon that could contribute to localized within-sample deviations from the broad trend linking coccolith mass to carbonate chemistry.

A general physiological mechanism is evidently insufficient to explain the overall decline in coccolith mass in our data set. The three noëlaerhabdacean genera include several species, each with a number of morphological variants (morphotypes)<sup>26</sup>. In our global sample set, the lightest *E. huxleyi* (morphotype C) was present in waters with  $[\text{CO}_3^{2-}] < 200 \mu\text{mol kg}^{-1}$ , whereas the heavy *Gephyrocapsa oceanica* (which includes several morphotypes<sup>13</sup>) only occurred in waters above this concentration (Fig. 1b). *Emiliania huxleyi* with intermediate mass (morphotypes A and B) occupied wider and intermediate  $[\text{CO}_3^{2-}]$  ranges. Significant coccolith malformation occurred at low  $[\text{CO}_3^{2-}]$  in *E. huxleyi* morphotype C from the Patagonian shelf and Antarctic

**Table 1 | Correlation between coccolith mass and physicochemistry**

Sample set	Temperature	Salinity	Alkalinity	DIC	pH	$p\text{CO}_2$	$\text{HCO}_3^-$	$\text{CO}_3^{2-}$	$\Omega\text{Ca}$	<i>n</i>
All data	<b>0.68</b>	<b>0.27</b>	<b>0.13</b>	<b>-0.82</b>	<b>0.75</b>	<b>-0.69</b>	<b>-0.87</b>	<b>0.86</b>	<b>0.86</b>	712
BIOSOPE	<b>0.77</b>	<b>0.58</b>	<b>0.63</b>	-0.02	<b>-0.52</b>	<b>0.53</b>	<b>-0.57</b>	<b>0.71</b>	<b>0.72</b>	84
PATAGONIA	0.31	0.18	0.14	0.03	0.04	0.01	-0.11	0.26	0.26	39
OISO-11	<b>0.65</b>	0.32	0.19	<b>-0.68</b>	-0.31	0.29	<b>-0.65</b>	<b>0.60</b>	<b>0.61</b>	57
MD90-0963	-0.12	<b>0.62</b>	<b>0.63</b>	<b>0.60</b>	0.37	-0.33	-0.07	<b>0.53</b>	<b>0.52</b>	68
MD97-2141	<b>-0.71</b>	-0.15	-0.12	<b>-0.68</b>	<b>0.79</b>	<b>-0.80</b>	<b>-0.75</b>	<b>0.67</b>	<b>0.67</b>	305
MD97-2138	-0.34	0.11	0.13	-0.41	<b>0.59</b>	<b>-0.60</b>	<b>-0.60</b>	<b>0.65</b>	<b>0.65</b>	35
MD02-2529	<b>-0.48</b>	<b>0.45</b>	<b>0.48</b>	0.23	<b>0.62</b>	<b>-0.62</b>	-0.23	<b>0.61</b>	<b>0.61</b>	124
MD94-103	-0.80	0.00	0.12	-0.50	<b>0.88</b>	<b>-0.87</b>	-0.75	0.64	0.67	9
SU90-08	-0.40	0.60	0.59	0.02	0.64	-0.65	-0.41	<b>0.68</b>	<b>0.68</b>	18

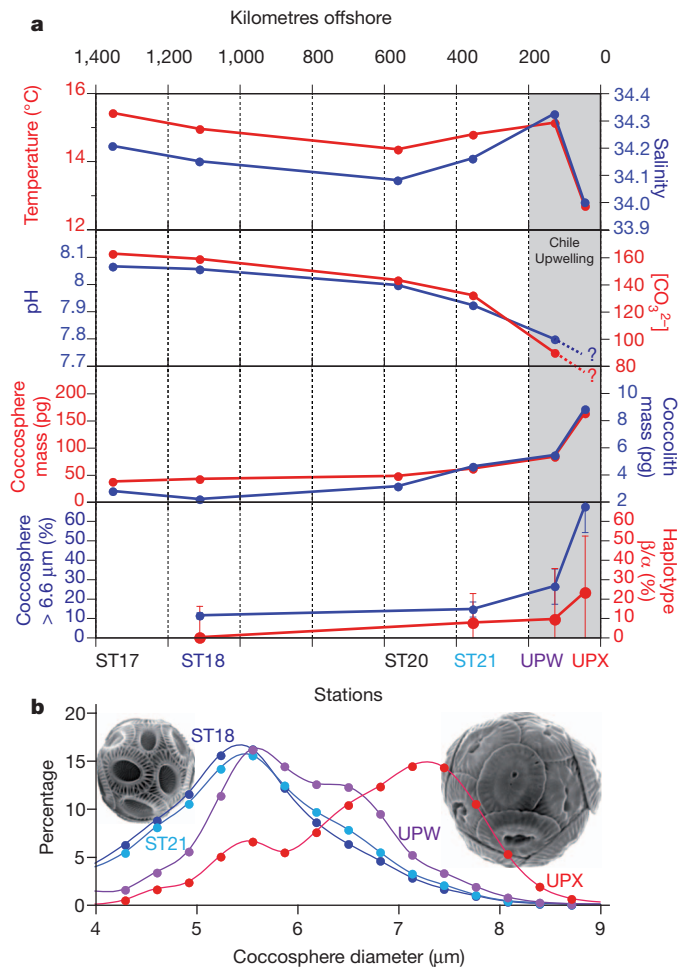
Significant correlation coefficient values are set in bold. *n*, number of samples in a given set.



**Figure 2 | Variation of coccolith mass, species composition and CO<sub>2</sub> concentration over the last 40 kyr.** **a**, Coccolith mass (filled circles) in the low-latitude (tropical) stack is shown in black (average of the four records, error bars are standard error between records). Coccolith mass at sample site MD94-103 (South Indian Ocean) is shown in red and at site SU90-08 (North Atlantic Ocean), in blue (error bars show standard error in each sample). The CO<sub>2</sub> concentration in ice cores at EPICA<sup>29</sup> (0–22 kyr) and Vostok<sup>30</sup> (22–40 kyr) is represented by crosses. **b**, Measurements of coccolith mass in the low-latitude stack, showing *G. oceanica* in blue, *E. huxleyi* in red and the relative abundance of the two taxa in black. Grey shading marks the period of the Last Glacial Maximum.

ocean. Changes in the relative abundance of taxa were therefore predominantly responsible for the decrease in coccolith mass with ocean acidification that was seen in modern samples. Both *Gephyrocapsa* and *Emiliania* showed a decrease in coccolith mass of about 25% from the Last Glacial Maximum to the near-present, paralleling an increase in CO<sub>2</sub> of about 100 parts per million by volume (Fig. 2). In addition to a physiological response of individual morphotypes, this could have resulted from changes in the abundance of different morphotypes within the genera. Superimposed on this intra-generic response, a decrease in relative abundance of *Gephyrocapsa* versus *Emiliania* was the main factor underlying the overall mass variation over the last 40 kyr (Fig. 2b). Although the function of coccoliths is unknown, our data indicate that variability in calcite mass (or the associated energy expenditure) is subject to ecological selective pressure.

We observed a key exception to the global correlation between noëlaerhabdacean calcification and [CO<sub>3</sub><sup>2-</sup>]. In Patagonian-shelf and Chilean upwelling waters with low [CO<sub>3</sub><sup>2-</sup>], in which the overall trend would predict low coccolith mass, we detected an unexpectedly highly calcified *E. huxleyi* morphotype (Fig. 1b), reminiscent of morphotype R<sup>26</sup>. The relative abundance of this morphotype increased with decreasing pH along the Pacific transect towards Chile (Fig. 3). Many environmental gradients exist along this transect, leading to the correlation of several factors with calcite mass, with mutual interactions (see for example ref. 18) that potentially mask the typical



**Figure 3 | Physico-chemical and coccolithophore variability along an east-west acidity gradient in the south-east Pacific.** Data are from the biogeochemistry and optics South Pacific experiment (BIOSOPE) (station averages of values are given in Supplementary Table 2). **a**, Plots of temperature and salinity, [CO<sub>3</sub><sup>2-</sup>] and pH (water chemistry was not measured at station UPX), mass of coccospheres and coccoliths, and percentages of large coccospheres (>6.6 μm) and of haplotype β versus α. Error bars are confidence intervals at 95%. **b**, Distribution of coccosphere diameters at different stations. ST18 (*n* = 1,334), ST21 (*n* = 578), UPW (*n* = 210), UPX (*n* = 203). Typical coccospheres of *E. huxleyi* type A (left, from ST18) and *E. huxleyi* type R (right, from UPX) are shown.

response of mass to carbonate chemistry. Alternatively, because coccolith morphotype is thought to be subject to genetic regulation<sup>11</sup>, this highly calcified *E. huxleyi* morphotype may be a genetic entity with an adaptation enabling it to calcify heavily in the relatively acidic upwelling waters. In this context, we note that the only culture study that reported an increase in *E. huxleyi* calcification with ocean acidification was conducted with a morphotype-R strain<sup>9</sup>.

To probe the genetic diversity of *E. huxleyi*, we generated clone libraries from samples collected along the acidity gradient that characterizes offshore Chilean water masses. We detected a shift in the relative abundance of two distinct mitochondrial haplotypes of *E. huxleyi* that coincided with the shift in relative abundance of morphotypes along the transect (Fig. 3). Each haplotype has wide oceanic distribution (Supplementary Information), indicating that the observed distribution is not the result of regional endemism. The relationship between mitochondrial haplotypes and morphotype is not, however, straightforward<sup>27</sup>, and establishing a strict link between morphotypic and genotypic diversities requires further environmental morphogenetic studies and/or culture-based physio-genomic comparisons. The presence of highly calcified *E. huxleyi* in these samples does not mask

the main pattern of decreasing calcification at low  $[\text{CO}_3^{2-}]$ , but highlights the fact that coccolithophores can calcify heavily at low pH (7.62) and low  $[\text{CO}_3^{2-}]$  (71  $\mu\text{mol kg}^{-1}$ ).

Coccolithophore calcification may be influenced by several factors, but our environmental data show a spatio-temporally consistent decline of coccolith mass with decreasing  $[\text{CO}_3^{2-}]$ . Integrating this coccolithophore response with the predicted decrease in calcification of planktonic foraminifera<sup>6,7</sup> and neritic corals<sup>3,4</sup> under conditions of increased  $\text{CO}_2$  means that entire marine calcifying communities seem likely to be affected in the future. However, the presence of highly calcified *E. huxleyi* in  $\text{CO}_2$ -rich modern waters demonstrates that prediction of future responses is unlikely to be straightforward. Such complexity could account for the lack of an obvious overall direction in the response of coccolithophore calcification over a potentially analogous ocean acidification event about 55 million years ago at the Palaeocene–Eocene Thermal Maximum<sup>28</sup>. Attention should now focus not only on the physiological response of individual strains to changing carbonate chemistry, but also on characterizing responses in complex assemblages.

## METHODS SUMMARY

Image analyses were performed on 40 frames ( $240 \times 180 \mu\text{m}$  with a pixel area of  $0.0225 \mu\text{m}^2$ ) per microscope slide, selected using an automated microscope (Leica DMRBE). Coccoliths and coccospheres were detected, classified and morphometrically analysed by the SYRACO software, which performs pattern recognition using artificial neural networks. Mass was estimated by measuring brightness in cross-polarized light (birefringence), with brightness being converted into mass after calibration with calcite microspheres of known mass.

Water sampling and measurements of temperature and salinity on three oceanographic cruises were conducted onboard. Total dissolved inorganic carbon (DIC) and total alkalinity (TA) were determined by the potentiometric acid titration method. To infer past changes in surface ocean-carbonate-system values, salinity and TA values in the past were estimated from geochemical proxies of past surface hydrography, using published estimates of past sea surface temperature and measured  $\delta^{18}\text{O}$ . Surface  $p_{\text{CO}_2}$  was estimated using atmospheric  $p_{\text{CO}_2}$  values measured in Antarctic ice cores at Vostok and EPICA, which are considered to represent global values throughout the temporal range of our records. Other carbonate-chemistry parameters were calculated using the CO2sys software with temperature, salinity, TA and DIC as inputs.

For genetic analyses, 10-litre seawater samples were filtered onto nucleopore membranes (Millipore). Total DNA was extracted using the DNeasy Plant mini kit (Qiagen) and *cox3* sequences were amplified using the Phusion High-Fidelity PCR mix (New England Biolabs). PCR primers are given in the Supplementary Information. For construction of environmental clone libraries, amplified sequences were inserted into the TOPO-TA PCR 4 vector (Invitrogen) before transformation into competent cells. PCR products were sequenced directly using the ABI PRISM BigDye Terminator Cycle Sequencing Kit on an ABI PRISM 3100 xl auto sequencer (Applied Biosystems). Maximum likelihood analysis was performed in TREEFINDER using manually aligned 812-base-pair *cox3* sequences under the corrected Akaike information criterion model.

Received 20 July 2010; accepted 15 June 2011.

1. Sabine, C. L. *et al.* The oceanic sink for anthropogenic  $\text{CO}_2$ . *Science* **305**, 367–371 (2004).
2. Fabry, V. J., Seibel, B. A., Feely, R. A. & Orr, J. C. Impacts of ocean acidification on marine fauna and ecosystem processes. *ICES J. Mar. Sci.* **65**, 414–432 (2008).
3. Gattuso, J.-P., Frankignoulle, M., Bourge, I., Romaine, S. & Buddemeier, R. W. Effect of calcium carbonate saturation of seawater on coral calcification. *Global Planet. Change* **18**, 37–46 (1998).
4. Kleypas, J. A. *et al.* Geochemical consequences of increased atmospheric carbon dioxide on coral reefs. *Science* **284**, 118–120 (1999).
5. Barker, S. & Elderfield, H. Foraminiferal calcification response to glacial–interglacial changes in atmospheric  $\text{CO}_2$ . *Science* **297**, 833–836 (2002).
6. Moy, A. D., Howard, W. R., Bray, S. G. & Trull, T. W. Reduced calcification in modern Southern Ocean planktonic foraminifera. *Nature Geosci.* **2**, 276–280 (2009).
7. de Moel, H. *et al.* Planktic foraminiferal shell thinning in the Arabian Sea due to anthropogenic ocean acidification? *Biogeosciences* **6**, 1917–1925 (2009).
8. Riebesell, U. *et al.* Reduced calcification of marine plankton in response to increased atmospheric  $\text{CO}_2$ . *Nature* **407**, 364–367 (2000).
9. Iglesias-Rodriguez, M. D. *et al.* Phytoplankton calcification in a high- $\text{CO}_2$  world. *Science* **320**, 336–340 (2008).
10. Langer, G. M. *et al.* Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochim. Geophys. Geosyst.* **7**, Q09006 (2006).

11. Langer, G., Nehrke, G., Probert, I., Ly, J. & Ziveri, P. Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosci. Discuss.* **6**, 4361–4383 (2009).
12. de Vargas, C., Aubry, M. P., Probert, I. & Young, J. in *Evolution of Aquatic Photoautotrophs* (eds Falkowski, P. G. & Knoll, A. H.) 251–285 (Academic, 2007).
13. Bollmann, J., Henderiks, J. & Brabec, B. Global calibration of *Gephyrocapsa* coccolith abundance in Holocene sediments for paleotemperature assessment. *Paleoceanogr.* **17**, doi:10.1029/2001PA000742 (2002).
14. Lewis, E. & Wallace, D. W. R. Program developed for  $\text{CO}_2$  system calculations. *Carbon Dioxide Information and Analysis Center Report, ORNL/CDIAC-105* (1998).
15. Hönisch, B. & Hemming, N. G. Surface ocean pH response to variations in  $p_{\text{CO}_2}$  through two full glacial cycles. *Earth Planet. Sci. Lett.* **236**, 305–314 (2005).
16. Paasche, E. A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification–photosynthesis interactions. *Phycologia* **40**, 503–529 (2001).
17. Zondervan, I. The effects of light, macronutrients, trace metals and  $\text{CO}_2$  on the production of calcium carbonate and organic carbon in coccolithophores—a review. *Deep Sea Res. II* **54**, 521–537 (2007).
18. Feng, Y. *et al.* Interactive effects of increased  $p_{\text{CO}_2}$ , temperature and irradiance on the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae). *Eur. J. Phycol.* **43**, 87–98 (2008).
19. Bollmann, J. & Herrle, J. O. Morphological variation of *Emiliania huxleyi* and sea surface salinity. *Earth Planet. Sci. Lett.* **255**, 273–288 (2007).
20. Colmenero-Hidalgo, E., Flores, J. A. & Sierro, F. J. Biometry of *Emiliania huxleyi* and its biostratigraphic significance in the Eastern North Atlantic Ocean and Western Mediterranean Sea in the last 20 000 years. *Mar. Micropaleontol.* **46**, 247–263 (2002).
21. Anderson, D. M. & Archer, D. Glacial interglacial stability of ocean pH inferred from foraminifer dissolution rates. *Nature* **416**, 70–73 (2002).
22. Buitenhuis, E. T., de Baar, H. J. W. & Veldhuis, M. J. W. Photosynthesis and calcification by *Emiliania huxleyi* (Prymnesiophyceae) as a function of inorganic carbon species. *J. Phycol.* **35**, 949–959 (1999).
23. Berry, L., Taylor, A. R., Lucken, U., Ryan, K. P. & Brownlee, C. Calcification and inorganic carbon acquisition in coccolithophores. *Funct. Plant Biol.* **29**, 289–299 (2002).
24. Mackinder, L., Wheeler, G., Schroeder, D., Riebesell, U. & Brownlee, C. Molecular mechanisms underlying calcification in coccolithophores. *Geomicrobiol. J.* **27**, 585–595 (2010).
25. Zondervan, I., Rost, B. & Riebesell, U. Effect of  $\text{CO}_2$  concentration on the PIC/POC ratio in the coccolithophore *Emiliania huxleyi* grown under light-limiting conditions and different daylengths. *J. Exp. Mar. Biol. Ecol.* **272**, 55–70 (2002).
26. Young, J. *et al.* A guide to extant coccolithophore taxonomy. *J. Nannoplankton Res.* **1** (Special Issue), 1–132 (2003).
27. Hagino, K. *et al.* New evidence for morphological and genetic variation in the cosmopolitan coccolithophore *Emiliania huxleyi* (Prymnesiophyceae) from the COX1b-ATP4 genes. *J. Phycol.* (in the press).
28. Gibbs, S. J., Bown, P. R., Sessa, J. A., Bralower, T. J. & Wilson, P. A. Nannoplankton extinction and origination across the Paleocene–Eocene Thermal Maximum. *Science* **314**, 1770–1773 (2006).
29. Monnin, E. *et al.* Atmospheric  $\text{CO}_2$  concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
30. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the crew from Puerto Deseado, Atalante, Suroit and Marion-Dufresne, and D. Vaulot, L. Garczarek, M.-A. Sicre and H. Claustre for their help in collecting material for this work. The long-term OISO observational programme is supported by INSU (Institut National des Sciences de l'Univers), IPSL (Institut Pierre-Simon Laplace) and IPEV (Institut Paul-Emile Victor). We thank F. C. Bassinot for help in estimating palaeosalinities. The IMAGES programme is acknowledged for collection and curation of the cores. This work was funded by the 'Agence National de la Recherche' project PALEO-CTD (grant ANR-06-JCJC-0142), by the European Research Council under grant agreement 205150, by the European Funding Agencies from the ERA-net program Biodiversa, under the Biomars project, and by the European Community's Seventh Framework Program EPOCA (European Project on Ocean Acidification) under grant agreement 211384.

**Author Contributions** On the basis of an original idea from L.B., the concept of this paper was developed in discussion between all authors. L.B., N.B., P.C. and M.G. conducted coccolith measurements, D.R.-P., N.M. and C.G. conducted modern-ocean chemistry measurements, L.B. and T.d.G.-T. computed past ocean chemistry, E.M.B., I.P. and C.d.V. performed genetic analyses, B.R., R.E.M.R. and I.P. conceptualized the physiological interpretation, L.B., I.P., D.R.-P., C.d.V. and R.E.M.R. interpreted the relationships between calcification and environment.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The GenBank accession numbers are JN098138–JN098158, JN098160 and JN098163–JN098174; their correspondence is given in the online Supplementary Information. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.B. ([beaufort@ceregr.fr](mailto:beaufort@ceregr.fr)).



# Excess digestive capacity in predators reflects a life of feast and famine

Jonathan B. Armstrong<sup>1</sup> & Daniel E. Schindler<sup>1</sup>

A central challenge for predators is achieving positive energy balance when prey are spatially and temporally heterogeneous. Ecological heterogeneity produces evolutionary trade-offs in the physiological design of predators; this is because the ability to capitalize on pulses of food abundance requires high capacity for food-processing, yet maintaining such capacity imposes energetic costs that are taxing during periods of food scarcity<sup>1,2</sup>. Recent advances in physiology show that when variation in foraging opportunities is predictable, animals may adjust energetic trade-offs by rapidly modulating their digestive system to track variation in foraging opportunities<sup>1</sup>. However, it is increasingly recognized that foraging opportunities for animals are unpredictable<sup>3</sup>, which should favour animals that maintain a capacity for food-processing that exceeds average levels of consumption (loads)<sup>2,4</sup>. Despite this basic principle of quantitative evolutionary design, estimates of digestive load:capacity ratios in wild animals are virtually non-existent<sup>1</sup>. Here we provide an extensive assessment of load:capacity ratios for the digestive systems of predators in the wild, compiling 639 estimates across 38 species of fish. We found that piscine predators typically maintain the physiological capacity to feed at daily rates 2–3 times higher than what they experience on average. A numerical simulation of the trade-off between food-processing capacity and metabolic cost suggests that the observed level of physiological opportunism is profitable only if predator–prey encounters, and thus predator energy budgets, are far more variable in nature than currently assumed.

Predation opportunities for animals in the wild are distributed heterogeneously in space and time owing to a variety of interactions between predator behaviour<sup>5</sup> (for example, how they allocate time to foraging, avoiding predators, and finding mates), prey behaviour<sup>6</sup>, and heterogeneity in the physical attributes of habitats where predator–prey interactions occur<sup>7,8</sup>. As a result, encounter rates between predators and prey in their natural habitats are exceedingly difficult to estimate<sup>6</sup> and may be far more heterogeneous than assumed. To cope with such ecological heterogeneity, predators employ behavioural and physiological tactics that allow them to store energy when food is plentiful, and utilize such reserves when food is scarce. For predators that store energy internally (for example, as fat reserves or somatic growth), physiological constraints on food-processing may limit rates of energy storage<sup>9</sup> and compensatory growth<sup>10</sup> and thus the potential to capitalize on pulses of food. To thrive in an environment where foraging opportunities vary widely but unpredictably, predators should maintain physiological opportunism and exhibit maximum capacities for food-processing that considerably exceed the average capacity required.

Acquiring energy from prey is a serial process that spans foraging (attacking, handling, and ingesting prey) to assimilation (digesting food and absorbing nutrients)<sup>11</sup>. In a homogeneous world, the most economic pathway would exhibit symmorphosis, where the maximal processing rate at each stage is equal to the average input rate (load) from the prior stage<sup>12</sup>. However, if the load at any stage varies in time, performance at that stage should exhibit excess capacity, such that the

maximal rate exceeds the mean and the system can accommodate spikes in load<sup>4</sup>. Surprisingly, the role of excess capacity along the food-to-fuel pathway is poorly understood. It is known that parts of the foregut (for example, the oesophagus, stomach, or other comparable structures) can act as food storage reservoirs, enabling predators to ingest prey faster than they can process it<sup>13,14</sup>. This excess capacity for feeding accommodates short-term variation in prey encounters<sup>14</sup> but it does not facilitate sustained increases in energy gain, which predators would require in order to build energy reserves or achieve compensatory growth. To sustain increased consumption rates, predators require excess capacity for assimilation (that is, digestion and absorption). Current knowledge of excess capacity in the digestive system is primarily limited to measurements at the cellular level<sup>2</sup>, and estimates of load:capacity ratios for integrated rates of assimilation are virtually non-existent for wild animals. Measurements of excess assimilation capacity are important for both physiologists, who seek to evaluate symmorphosis (or lack of it) across the energy intake chain, and ecologists, who lack reliable data to characterize the variation in foraging opportunities experienced by predators in the wild. Here we quantitatively assess how daily variation in foraging opportunity should affect the profitability of excess assimilative capacity, and use our results to interpret observed load:capacity ratios from animals in the wild.

Predators require excess assimilative capacity to capitalize on large pulses of prey abundance as reflected in the positive tail of the distribution of predator foraging opportunities<sup>8,14</sup>. However, increasing assimilative capacity also increases the energetic costs of physiological maintenance and locomotion<sup>1</sup>. We developed a simple numerical simulation to characterize this trade-off and explore its response to the underlying distribution describing daily variation in foraging opportunity. We simulated variation in daily foraging opportunities using a gamma distribution with a fixed arbitrary value for the mean, but different coefficients of variation (c.v., defined as (s.d./mean) × 100) to simulate different levels of variability in the environment (see Methods, Supplementary Fig. 1). A type I functional response<sup>15</sup> modelled how assimilative capacity ( $C_{\max}$ , the maximum amount of food that can be consumed in one day) determines the daily food consumption,  $C$ , that a predator derives from a daily foraging opportunity,  $x$ :

$$C(x) = \begin{cases} x & \text{if } x < C_{\max} \\ C_{\max} & \text{if } x \geq C_{\max} \end{cases} \quad (1)$$

The expected consumption rate ( $J d^{-1}$ ) for an individual with a specified  $C_{\max}$  and gamma-distributed variation in daily foraging opportunity is:

$$E(C_{\max}) = \int_0^{\infty} C(x)g(x|k, \theta)dx \quad (2)$$

where  $C(x)$  is the type I functional response of equation (1), and  $g(x)$  is the gamma distribution with shape and scale parameters ( $k$  and  $\theta$ ) varied to produce specified levels of mean and c.v.

<sup>1</sup>School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, Washington 98195, USA.

The load:capacity ratio for assimilation is given by equation (3) below, and its inverse is synonymous with excess capacity,  $Z$ :

$$E(C_{\max})/C_{\max} = 1/Z \quad (3)$$

The net energetic profit ( $N$ ) is the energy remaining after gains from consumption are paid to maintenance and activity costs,  $M$ :

$$N = E(C_{\max}) - M \quad (4)$$

The energetic costs associated with increased assimilative capacity are poorly described and are likely to vary among taxa. We model  $M$  as a constant fraction,  $r$ , of  $C_{\max}$  (see Methods):

$$M = rC_{\max}, \text{ where } 0 < r < 1 \quad (5)$$

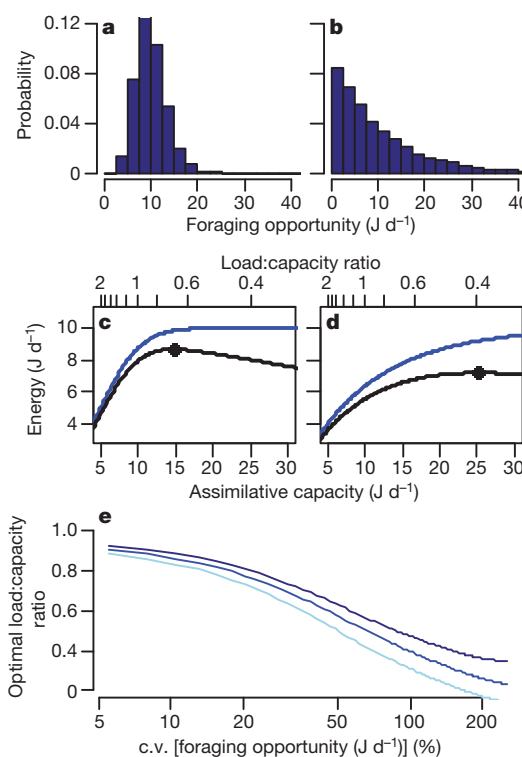
thus

$$N = E(C_{\max}) - rC_{\max} \quad (6)$$

In our analysis, we used numerical simulation to calculate the load:capacity ratio that maximized the net profit function (equation (6)), given the c.v. in daily foraging opportunity and the cost of the gut. The optimal load:capacity ratio for assimilation decreased with increasing variation in foraging opportunity, but did not exhibit substantial levels of excess capacity until the c.v. was very high, indicating strong right-skew in the distribution of foraging opportunities (Fig. 1 and Supplementary Fig. 1). For example, the optimal load:capacity ratio reached  $\sim 0.4$  when the c.v. in foraging opportunity ranged from 70% to 150%, depending on the costs of maintaining excess capacity (Fig. 1e). This suggests that, in order for integrated rates of assimilation to exhibit levels of excess capacity seen in many other biological structures<sup>2</sup> (for example, load:capacity ratios  $< 0.5$ ), predator–prey encounters would need to be extremely heterogeneous in time.

To assess the variation in foraging opportunities that predators experience in nature, we compared our model scenarios to assimilative load:capacity ratios estimated for piscine predators in the wild. We compiled 639 estimates of mean daily consumption rate in fishes. The estimates were calculated from bioenergetics models that integrate across field measurements of growth and express consumption relative to maximum physiological rates. This bioenergetics framework<sup>16,17</sup> (described in Supplementary Information) has been parameterized for different species through laboratory studies that measure the effects of water temperature and body mass on both metabolism and  $C_{\max}$ , the daily consumption rate observed under *ad libitum* feeding.  $C_{\max}$  is determined by assimilative capacity<sup>18</sup>, and corresponds well to the consumption rates of wild fish in conditions where food is unlimited<sup>19</sup>. The model uses field data on predator growth achieved over a defined time period, diet composition and quality, and water temperature to estimate the proportion of  $C_{\max}$  (which we call  $p$ ) that balances an energy budget where growth is surplus energy after energetic gains from consumption are paid to metabolism, excretion and specific dynamic action. Thus,  $p$  represents the load:capacity ratio for assimilation. We searched research databases for all papers citing this bioenergetics model and compiled all estimates of  $p$  recorded in the wild (Supplementary Table 1). We grouped  $p$  estimates by population (unique combinations of species and publication,  $n = 66$ ). The mean length of time over which  $p$  was estimated was 208 days (s.d., 143 days).

Across 66 populations from 38 species, the median  $p$  was 43% (s.d., 16%), indicating a load:capacity ratio for assimilation of 0.43. The distribution of  $p$  was right-skewed and similar in shape whether grouped by population or not (Fig. 2). Populations from marine ( $n = 11$ ), lake ( $n = 49$ ) and stream ( $n = 6$ ) environments did not exhibit significantly different load:capacity ratios (Kruskal–Wallis test:  $P = 0.10$ , d.f. = 2) and less than 5% of populations exhibited load:capacity ratios greater than 0.8. Among the  $>60\%$  of populations that

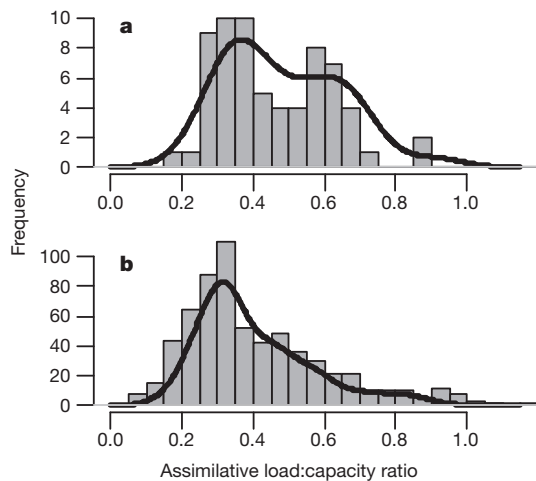


**Figure 1 | Results from a simulation model exploring the energetic profitability of excess capacity for assimilation as a function of the daily variation in foraging opportunity.** We define assimilation as digestion and absorption. **a, b**, Gamma distributions characterizing two scenarios of ecological heterogeneity that both yield an average of 10 energy units per day in foraging opportunity. **a**, Variance in daily foraging opportunity is equal to the mean, as from a Poisson distribution (c.v. = 32%,  $\theta = 1$ ,  $k = 10$ ). See main text for nomenclature. **b**, Foraging opportunities are highly right-skewed and show 10 times more variance than a Poisson distribution with similar mean (c.v. = 100%,  $\theta = 10$ ,  $k = 1$ ). **c, d**, Cost–benefit analysis of excess assimilative capacity under the two scenarios of ecological heterogeneity; **c** and **d** correspond respectively to distributions displayed in **a** and **b**. Lines represent gross and net energetic gains (blue and black, respectively) resulting from different levels of assimilative capacity (that is, maximum daily consumption rate). The upper x-axis measures excess capacity in terms of the load:capacity ratio (the mean daily consumption rate relative to the maximum rate). Filled circle represents the energetically optimal digestive capacity. Results shown are from the medium cost scenario (see Supplementary Information for a full description). **e**, Model results showing the energetically optimal load:capacity ratio as a function of the c.v. in daily foraging opportunity. Curves depict different cost scenarios (light blue, low; blue, medium; dark blue, high: see Supplementary Information for full description).

exhibited load: capacity ratios less than 0.5, all functional guilds of predators were represented, including planktivores, benthivores and piscivores.

Digestive machinery is expensive<sup>1,20</sup>, so why do piscine predators maintain the physiological capacity to feed at daily rates that are 2–3 times higher than what they achieve on average? Our numerical simulation demonstrates that such low ratios of assimilative load: capacity become energetically profitable when the distribution of daily foraging opportunities is highly heterogeneous (that is, right-skewed; Fig. 1, Supplementary Fig. 1). This suggests that episodes of gorging and fasting are common in fishes, and occur not only in ambush predators that pursue large prey<sup>21</sup>, but also in predators that feed on insects and zooplankton. These results question common assumptions regarding the variance in foraging opportunities experienced by predators in the wild.

Foraging models frequently simulate variation in a predator's feeding opportunities by drawing prey encounters from the Poisson



**Figure 2 | Histograms showing bioenergetics estimates of integrated consumption rates in wild piscine predators, expressed as the assimilative load:capacity ratio.** This ratio is defined as average daily consumption rate divided by the physiological maximum. **a**, Estimates grouped by unique combinations of species and publication ( $n = 66$ ). **b**, All estimates ( $n = 639$ ). Lines show a non-parametric density estimate of each distribution (see Methods).

distribution<sup>21,22</sup>, which exhibits equal mean and variance ( $\lambda$ ). Under most configurations of the Poisson (for example, if  $\lambda \gtrsim 3$ ), values greater than twice the mean have extremely low probabilities (0–3% of occurrences). Thus, foraging models frequently simulate a world in which predators would almost never take advantage of the excess assimilative capacity that we have shown in fishes (that is, median capacity is 2.3 times expected load). This suggests that the feeding opportunities of piscine predators, and perhaps many others, are more heterogeneous than has been appreciated in ecology.

In the past decade, three meta-analyses have concluded that lizards<sup>23</sup>, fish<sup>24</sup>, birds and mammals<sup>25</sup> rarely exhibit negative energy budgets at daily timescales, based on observations that predators rarely have completely empty stomachs<sup>23,24</sup>, or that their integrated rates of energy gain exceed losses at seasonal to annual timescales<sup>25</sup>. In our analysis, we found that piscine predators probably experience foraging opportunities that are distributed extremely patchily in time, such that the median population would actually spend 18–53% of days feeding at levels that are lower than a typical maintenance ration (~20% of  $C_{\max}$ ; ref. 17). This suggests that fishes routinely run negative energy budgets and warns that prior syntheses of empirical data may have underestimated variability in the daily energy balance of not only fishes, but other taxa as well. Future work should consider how patterns of capacity through the food–fuel pathway are associated with different strategies of energy regulation. For example, many small birds and mammals hoard food during resource pulses or employ torpor to avoid energy deficits<sup>26</sup>. These mechanisms may stabilize consumption rates and energy budgets enough to make maintaining excess capacity for assimilating food unnecessary, but would probably require increased capacity for handling food.

We have assumed that the assimilative capacity observed in fishes has evolved to optimize energy budgets under trade-offs between the capacity and metabolic cost of the gut. There is widespread evidence that the energetic profitability of the gut has strong fitness consequences<sup>1,20</sup> and that its morphology and function evolve in response to energetic demands<sup>27</sup>. Further, the fishes in our analyses exhibit indeterminate somatic growth that is indeed limited by energy acquisition<sup>28</sup>. Therefore it is reasonable to use an energy maximization model to interpret the function of physiological capacity. The temporal pattern of foraging opportunities may vary among generations such that the optimal level of digestive capacity is not static in time. Our general conclusions are robust to such evolutionary disequilibrium because we found high levels

of excess digestive capacity across numerous taxa inhabiting diverse environments. The most parsimonious explanation of these findings is that piscine predators have evolved excess assimilative capacity to profit in a world where binging and fasting are regular occurrences.

Phenotypic flexibility can improve performance if energy budgets vary predictably<sup>1</sup>. For example, female mammals increase digestive capacity during pregnancy, and birds adjust digestive capacity during migration as they alternate between fuelling and flight<sup>1</sup>. Although seasonal changes in food abundance may be predictable, pulses of food are frequently both unpredictable and ephemeral<sup>29</sup>, causing the performance of consumers to be measured by the guts they possess in the moment and not the ones they can construct days or weeks later (except certain sit-and-wait predators that are capable of rapid intestinal regulation<sup>20</sup>). Similarly to the fishes in our analysis, birds may maintain excess capacity for consumption despite their ability to modify gut size over relatively short time-scales<sup>30</sup>. Flexibility in gut size has been suggested<sup>30</sup> to facilitate adaptive responses to long-term but not short-term variation in resource abundance. Our meta-analysis and model strongly suggest that predator–prey encounters are patchier than most ecologists assume, that fish run negative energy budgets more frequently than assumed, and that eco-physiology should consider the combined roles of excess capacity and phenotypic flexibility when considering how organisms cope with a world far more heterogeneous than the laboratory or treadmill. The insights derived from our analysis have important implications for models of predator–prey dynamics, community structure and the stability of food webs, which can be sensitive to heterogeneity in predator–prey interactions<sup>6</sup>.

## METHODS SUMMARY

To characterize the assimilative load:capacity ratios of piscine predators, we compiled data from bioenergetics analyses that express integrated consumption rates relative to the physiological maximum. These models<sup>16,17</sup> use field measurements of growth at monthly to annual timescales to reconstruct consumption rates, based on the energy budget:

$$G = C - (R + W + \text{SDA}) \quad (7)$$

where  $G$  is growth,  $C$  is food consumption,  $R$  is active metabolism,  $W$  is waste (faeces and urine) and SDA is specific dynamic action (the cost of digestion). Empirically derived, species-specific functions model the effects of water temperature and body mass on respiration and the maximum daily consumption rate ( $C_{\max}$ ). Additional functions model waste and SDA as taxes on consumption. To incorporate thermal and allometric constraints on  $C$ , the parameter  $p$  scales the actual consumption rate relative to the maximum consumption rate according to  $p = C/C_{\max}$ . Thus, equation (7) becomes:

$$G = pC_{\max} - (R + W + \text{SDA}) \quad (8)$$

On the basis of observed growth, diet composition, predator and prey energy densities, and the temperature regime experienced by a fish for the time interval over which growth is estimated, the model solves for the value of  $p$  that produces the value of  $G$  observed in the field. Because  $C_{\max}$  is determined by rates of food-processing<sup>18</sup>,  $p$  represents the load:capacity ratio for assimilation—the process that includes both the break down of food macromolecules and the transportation of nutrients across the gut wall.

For our meta-analysis of  $p$  estimates, we searched Web of Science and Google Scholar for all publications ( $n = 345$ ) citing the bioenergetics framework outlined above<sup>16,17</sup>. We compiled data from all publications (42 journal articles and 4 theses) that studied predators *in situ* and reported the parameter  $p$ . The metadata and source publications are provided in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 March; accepted 27 May 2011.**

**Published online 6 July 2011.**

1. Piersma, T. & van Gils, J. A. *The Flexible Phenotype* (Oxford Univ. Press, 2011).
2. Diamond, J. Quantitative evolutionary design. *J. Physiol. (Lond.)* **542**, 337–345 (2002).



3. Humphries, N. E. *et al.* Environmental context explains Levy and Brownian movement patterns of marine predators. *Nature* **465**, 1066–1069 (2010).
4. Gans, C. Momentarily excessive construction as the basis for protoadaptation. *Evolution* **33**, 227–233 (1979).
5. Ritchie, E. G. & Johnson, C. N. Predator interactions, mesopredator release and biodiversity conservation. *Ecol. Lett.* **12**, 982–998 (2009).
6. Walters, C. J. M. & Martell, S. J. D. *Fisheries Ecology and Management* (Princeton Univ. Press, 2004).
7. Kauffman, M. J. *et al.* Landscape heterogeneity shapes predation in a newly restored predator-prey system. *Ecol. Lett.* **10**, 690–700 (2007).
8. Armstrong, J. B. *et al.* Thermal heterogeneity mediates the effects of pulsed subsidies across a landscape. *Ecology* **91**, 1445–1454 (2010).
9. Klaassen, M., Lindstrom, A. & Zijlstra, R. Composition of fuel stores and digestive limitations to fuel deposition rate in the long-distance migratory thrush nightingale, *Luscinia luscinia*. *Physiol. Zool.* **70**, 125–133 (1997).
10. Dupont-Prinet, A. *et al.* Physiological mechanisms underlying a trade-off between growth rate and tolerance of feed deprivation in the European sea bass (*Dicentrarchus labrax*). *J. Exp. Biol.* **213**, 1143–1152 (2010).
11. Weiner, J. Physiological limits to sustainable energy budgets in birds and mammals: ecological implications. *Trends Ecol. Evol.* **7**, 384–388 (1992).
12. Weibel, E. R., Taylor, C. R. & Hoppeler, H. The concept of symmorphosis: a testable hypothesis of structure-function relationship. *Proc. Natl Acad. Sci. USA* **88**, 10357–10361 (1991).
13. Kersten, M. & Visser, W. The rate of food processing in the oystercatcher: food intake and energy expenditure constrained by a digestive bottleneck. *Funct. Ecol.* **10**, 440–448 (1996).
14. Essington, T. E., Hodgson, J. R. & Kitchell, J. F. Role of satiation in the functional response of a piscivore, largemouth bass (*Micropterus salmoides*). *Can. J. Fish. Aquat. Sci.* **57**, 548–556 (2000).
15. Holling, C. S. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *Can. Entomol.* **91**, 293–320 (1959).
16. Kitchell, J. F., Stewart, D. J. & Weininger, D. Applications of a bioenergetics model to yellow perch (*Perca flavescens*) and walleye (*Stizostedion vitreum-vitreum*). *J. Fish. Res. Board Can.* **34**, 1922–1935 (1977).
17. Fish Bioenergetics 3.0 (University of Wisconsin-Madison Centre for Limnology/Wisconsin Sea Grant Institute, 1997).
18. Elliott, J. M. & Persson, L. Estimation of daily rates of food consumption for fish. *J. Anim. Ecol.* **47**, 977–991 (1978).
19. Kitchell, J. F. *et al.* Predator-prey dynamics in an ecosystem context. *J. Fish Biol.* **45**, 209–226 (1994).
20. Secor, S. M., Stein, E. D. & Diamond, J. Rapid up-regulation of snake intestine in response to feeding: a new model of intestinal adaptation. *Am. J. Physiol.* **266**, G695–G705 (1994).
21. Breck, J. E. Foraging theory and piscivorous fish: are forage fish just big zooplankton? *Trans. Am. Fish. Soc.* **122**, 902–911 (1993).
22. DeAngelis, D. L. & Gross, L. J. *Individual-based Models and Approaches in Ecology. Populations, Communities, and Ecosystems* (Chapman and Hall, 1992).
23. Huey, R. B., Pianka, E. R. & Vitt, L. J. How often do lizards “run on empty”? *Ecology* **82**, 1–7 (2001).
24. Arrington, D. A., Winemiller, K. O., Loftus, W. F. & Akin, S. How often do fishes “run on empty”? *Ecology* **83**, 2145–2151 (2002).
25. Jeschke, J. M. When carnivores are “full and lazy”. *Oecologia* **152**, 357–364 (2007).
26. Brodin, A. & Clark, C. in *Foraging* (eds Stephen, D. S., Brown, J. S. & Ydenberg, R. C.) 221–269 (Univ. Chicago Press, 2007).
27. Hofmann, R. R. Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. *Oecologia* **78**, 443–457 (1989).
28. Schindler, D. E. & Eby, L. A. Stoichiometry of fishes and their prey: implications for nutrient recycling. *Ecology* **78**, 1816–1831 (1997).
29. Yang, L. H., Bastow, J. L., Spence, K. O. & Wright, A. N. What can we learn from resource pulses? *Ecology* **89**, 621–634 (2008).
30. McWilliams, S. R. & Karasov, W. H. Phenotypic flexibility in digestive system structure and function in migratory birds and its ecological significance. *Comp. Biochem. Physiol. A* **128**, 577–593 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Essington, R. Huey, T. Reed, A. Walters, V. Sturtevant and A. Armstrong for comments on this manuscript. We also thank the following people for contributing to this project: J. Kitchell, O. Jensen, E. Ward, D. Beauchamp, B. Chasco, A. Farrell, P. Bisson and J. Kershner. This work was supported by the Gordon and Betty Moore Foundation, the US National Science Foundation and the University of Washington School of Aquatic and Fishery Sciences.

**Author Contributions** J.B.A. and D.E.S. contributed to each stage of the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.B.A. (Jonny99@uw.edu).

## METHODS

**The fish bioenergetics model (FBEM).** This model<sup>16,17</sup> is used to estimate assimilative load:capacity ratios. It uses an energy balance approach to bioenergetics that provides an adaptable modelling framework, which has been applied to a wide variety of species and ecosystems. The model balances an energy budget where inputs must equal outputs, specifically:

$$C = G + R + W + \text{SDA} \quad (9)$$

Here  $C$  is food consumption,  $G$  is growth,  $R$  is active metabolism,  $W$  is waste (faeces and urine), and SDA is specific dynamic action (the cost of digestion). The FBEM provides functions that model the effects of water temperature and body mass on respiration and the maximum daily consumption rate ( $C_{\max}$ ). These functions are parameterized for individual species through controlled laboratory experiments. To formulate the respiration functions, oxygen consumption is measured in experiments that manipulate body size and water temperature. In addition, an activity multiplier is calculated to relate basal and active metabolism. To formulate the  $C_{\max}$  functions, daily consumption rates are measured in *ad libitum* feeding experiments that manipulate temperature and body size. Additional functions model waste and SDA as taxes on consumption. A full description of these functions is available in the Bioenergetics 3.0 software manual<sup>17</sup>.

Researchers typically use the FBEM to estimate consumption rates based on the observed growth rates of fishes in the field. This is done by rearranging the original terms in the energy budget so that growth represents surplus energy after gains from consumption are paid to energetic costs:

$$G = C - (R + W + \text{SDA}) \quad (10)$$

Growth is measured directly over some relatively long time interval (for example, monthly to annual growth increments) and the model is used to estimate the consumption rate  $C$  required to satisfy equation (10) to produce the observed growth increment. To incorporate thermal and allometric constraints on  $C$ , the parameter  $p$  scales the actual consumption rate relative to the maximum consumption rate according to  $p = C/C_{\max}$ . Thus, equation (10) becomes:

$$G = pC_{\max} - (R + W + \text{SDA}) \quad (11)$$

On the basis of observed growth, diet composition, predator and prey energy densities, and the temperature regime experienced by a fish for the time interval over which growth is estimated, the model solves for the value of  $p$  that produces the value of  $G$  observed in the field. The model is implemented at a daily time step and linearly interpolates between observations of temperature, diet composition and energy density. Because growth of fishes is indeterminate, the value of  $p$  is a sensitive but informative integrated parameter that reflects the realized consumption rate relative to the physiological maximum rate that an individual fish is capable of, based on its metabolic parameters, its body size, its energy density and the energy density of its prey, and water temperature. Because  $C_{\max}$  is determined by rates of food-processing<sup>18</sup>,  $p$  represents the load:capacity ratio for assimilation—the process that includes both the break-down of food macromolecules and the transportation of nutrients across the gut wall. Consumption estimates from the FBEM have been independently corroborated by intensive field-based methods<sup>31,32</sup>.

For our meta-analysis of  $p$  estimates, we searched Web of Science and Google Scholar for all publications ( $n = 345$ ) citing the bioenergetics framework outlined above<sup>16,17</sup>. We compiled data from all publications that studied predators *in situ* and reported the parameter  $p$  (42 journal articles and 4 theses). The meta-data and source publications are provided in Supplementary Table 1. To characterize distributions of  $p$  estimates (Fig. 2), we fitted kernel density estimates to the data in R<sup>33</sup>. We used a Gaussian smoothing kernel and selected the bandwidth following standard methods<sup>34</sup>.

**Modelling trade-offs associated with digestive capacity.** Here we give details of our numerical model in which we simulate these trade-offs. We used the gamma distribution to simulate variation in daily foraging opportunity because it can generate a diversity of biologically realistic shapes, ranging from narrow and

Gaussian to extremely right-skewed, representing an environment where prey encounters are extremely patchy. The gamma distribution has two parameters, namely  $k$  and  $\theta$ , which relate to the mean and variance of the distribution as follows:

$$\begin{aligned} \bar{x} &= k\theta \\ \sigma^2 &= k\theta^2 \end{aligned} \quad (12)$$

Rearranging equation (12) yields:

$$\begin{aligned} k &= \bar{x}^2/\sigma^2 \\ \theta &= \sigma^2/\bar{x} \end{aligned} \quad (13)$$

In order to examine the effect of variability in foraging opportunity on the energetic profitability of excess assimilative capacity, we varied  $\theta$  and  $k$  to generate gamma distributions with an equal mean, but different levels of variation. This simulates ecosystems that have an equal amount of prey available to predators, but different levels of variation in the temporal patterning of daily predator–prey encounters.

To provide biologically interpretable results, we used the coefficient of variation (c.v., defined as (s.d./mean)  $\times$  100) to describe variation in the gamma distribution. Unlike other metrics of variation (for example, the variance to mean ratio) the relationship between the c.v. and the shape of the gamma distribution scales isometrically with the mean of the distribution (for example, if the c.v. is held constant, changing the mean does not affect the skew of the distribution). Therefore, we could explore all ecologically relevant shapes of the gamma distribution by changing the c.v. of the distribution (range, 0.1–500%) while keeping the mean fixed at an arbitrary value. Supplementary Fig. 1 shows a subset of the gamma distributions that were generated in our simulations to illustrate the diversity of shapes that we considered.

The relationship between the maintenance cost and assimilative capacity of the gut ( $R_{\text{gut}}$  and  $C_{\max}$ ) is poorly documented. For the sake of parsimony, and because empirical data suggest that assimilative capacity in fish is related to the surface area of the gut<sup>35</sup>, we modelled a linear relationship between  $C_{\max}$  and  $R_{\text{gut}}$ , such that a proportional change in  $C_{\max}$  produces the same proportional change in  $R_{\text{gut}}$  (for example, doubling  $C_{\max}$  would double  $R_{\text{gut}}$ ). The relationship between  $C_{\max}$  and total metabolic cost,  $M$ , depends on the ratios of  $R_{\text{gut}}:M$  and  $M:C_{\max}$ . Under the assumed linear relationship, the slope ( $r$ ) of  $M$  as a function of  $C_{\max}$  is:

$$r = R_{\text{gut}}/R_{\text{tot}} \times R_{\text{tot}}/C_{\max} \quad (14)$$

We assumed that one-third of total daily respiration goes to maintaining digestive capacity, as this is the approximate proportion of total cardiac output delivered to an empty gut<sup>36</sup>. The empirically derived functions in the FBEM<sup>17</sup> suggest energetic losses due to respiration typically represent one-third to one-ninth of  $C_{\max}$  in fishes. Given these estimations,  $r$  ranges from 0.04 to 0.11. Because the  $y$ -intercept of a linear cost function (here, the cost of metabolism not associated with the gut) does not affect the optimal value for profit maximization, we modelled total metabolic cost,  $M$ , as a constant fraction,  $r$ , of  $C_{\max}$ , and included three cost scenarios: low ( $r = 4\%$ ), medium ( $r = 8\%$ ) and high ( $r = 12\%$ ). We did not consider specific dynamic action or excretion in our cost analysis, because we assumed these were taxes on consumption that do not represent physiological maintenance.

1. Rice, J. A. & Cochran, P. A. Independent evaluation of a bioenergetics model for largemouth bass. *Ecology* **65**, 732–739 (1984).
2. Beauchamp, D. A., Stewart, D. J. & Thomas, G. L. corroboration of a bioenergetics model for sockeye salmon. *Trans. Am. Fish. Soc.* **118**, 597–607 (1989).
3. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna); available at (<http://www.R-project.org>) (2010).
4. Scott, D. W. *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, 1992).
5. Stevens, E. D. & Devlin, R. H. Intestinal morphology in growth hormone transgenic coho salmon. *J. Fish Biol.* **56**, 191–195 (2000).
6. Farrell, A. P. et al. Gut blood flow in fish during exercise and severe hypercapnia. *Comp. Biochem. Physiol. A* **128**, 549–561 (2001).

# Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats

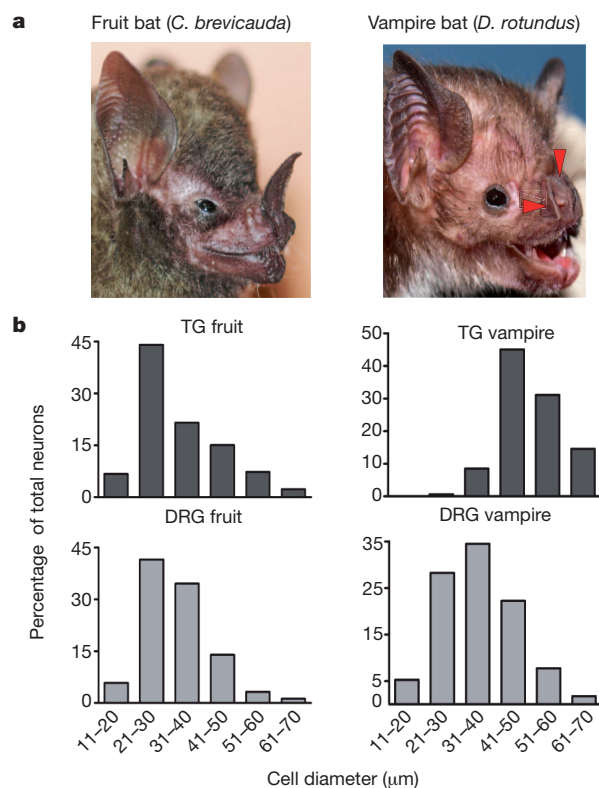
Elena O. Gracheva<sup>1\*</sup>, Julio F. Cordero-Morales<sup>1\*</sup>, José A. González-Carcacia<sup>2</sup>, Nicholas T. Ingolia<sup>3</sup>, Carlo Manno<sup>4</sup>, Carla I. Aranguren<sup>2</sup>, Jonathan S. Weissman<sup>5,6,7</sup> & David Julius<sup>1,5</sup>

Vampire bats (*Desmodus rotundus*) are obligate blood feeders that have evolved specialized systems to suit their sanguinary lifestyle<sup>1–3</sup>. Chief among such adaptations is the ability to detect infrared radiation as a means of locating hotspots on warm-blooded prey. Among vertebrates, only vampire bats, boas, pythons and pit vipers are capable of detecting infrared radiation<sup>1,4</sup>. In each case, infrared signals are detected by trigeminal nerve fibres that innervate specialized pit organs on the animal's face<sup>5–10</sup>. Thus, vampire bats and snakes have taken thermosensation to the extreme by developing specialized systems for detecting infrared radiation. As such, these creatures provide a window into the molecular and genetic mechanisms underlying evolutionary tuning of thermoreceptors in a species-specific or cell-type-specific manner. Previously, we have shown that snakes co-opt a non-heat-sensitive channel, vertebrate TRPA1 (transient receptor potential cation channel A1), to produce an infrared detector<sup>6</sup>. Here we show that vampire bats tune a channel that is already heat-sensitive, TRPV1, by lowering its thermal activation threshold to about 30 °C. This is achieved through alternative splicing of TRPV1 transcripts to produce a channel with a truncated carboxy-terminal cytoplasmic domain. These splicing events occur exclusively in trigeminal ganglia, and not in dorsal root ganglia, thereby maintaining a role for TRPV1 as a detector of noxious heat in somatic afferents. This reflects a unique organization of the bat *Trpv1* gene that we show to be characteristic of Laurasiatheria mammals (cows, dogs and moles), supporting a close phylogenetic relationship with bats. These findings reveal a novel molecular mechanism for physiological tuning of thermosensory nerve fibres.

Vampire bats have three 'leaf pits' that surround the nose (Fig. 1a) and receive input from low-threshold heat-sensitive nerve fibres responding to stimuli at temperatures >29 °C (refs 1, 7, 11, 12). Closely related fruit bats (*Carollia brevicauda*) have a different nasal structure devoid of pit organs (Fig. 1a) and thus cannot detect infrared stimuli<sup>3</sup>. Sensory ganglia of fruit bats showed a typical size distribution of neurons resembling that seen in other mammals (Fig. 1b). In contrast, trigeminal ganglia (TG) from vampire bats showed marked skewing towards large diameter neurons, much like that observed in TG of pit-bearing snakes<sup>6</sup>, suggesting that anatomical specialization of TG in vampire bats has similarly evolved to suit a predominant role in infrared sensation. Consistent with this, vampire bat dorsal root ganglia (DRG) showed a normal size distribution, resembling sensory ganglia of fruit bats or other mammals (Fig. 1b). To our knowledge, vampire bats provide the only such example of TG specialization among mammalian species.

Given the distinct anatomy of vampire bat TG versus DRG, we asked whether these ganglia show differential patterns of gene expression that might highlight molecules specifically involved in infrared sensation, as in the case of infrared-sensing snakes<sup>6</sup>. Thus, we performed 'deep sequencing' of complementary DNAs from bat TG and DRG.

Initially, transcriptomes from these ganglia appeared to be indistinguishable (Supplementary Fig. 1). However, examination of cDNAs encoding candidate thermosensors revealed a novel short isoform of the capsaicin receptor, TRPV1, an excitatory ion channel that is activated by noxious heat (>43 °C) (refs 13–15). This isoform (TRPV1-S) lacks 62 amino acids from the carboxy terminus (Fig. 2a and Supplementary Fig. 2) and is uniquely expressed in TG from vampire bat, constituting between 35% and 46% of all TRPV1 transcripts in TG, but <3% in DRG, as determined from transcriptome data or direct cloning and sequencing of RT-PCR (PCR of reverse transcripts) products amplified from sensory ganglia (Fig. 2b). Moreover, TRPV1-S transcripts were barely detected (<6%) in fruit bat TG or DRG

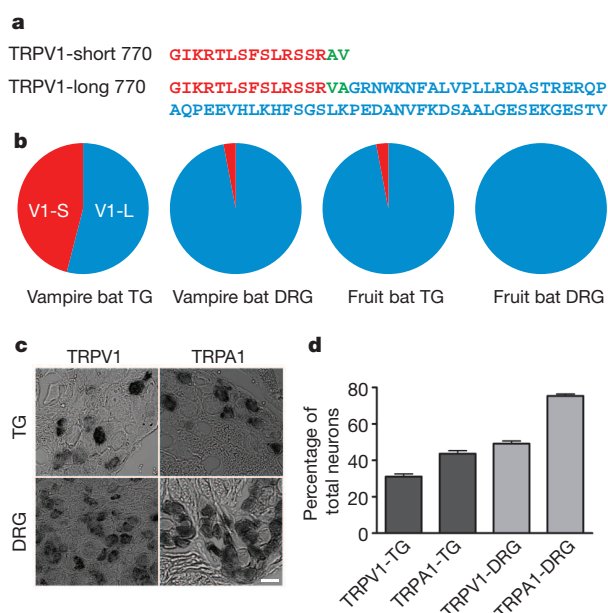


**Figure 1 | Anatomy of fruit bat and vampire bat sensory ganglia.** **a**, Facial anatomy of fruit bat (left) and vampire bat (right). Red arrowheads mark pit organs surrounding the vampire bat's nose. **b**, Neuronal cell sizes determined from histological sections of fruit bat TG ( $n = 300$ , 7 independent sections), fruit bat DRG ( $n = 345$ , 9 sections), vampire bat TG ( $n = 164$ , 6 sections) and vampire bat DRG ( $n = 400$ , 11 sections).

<sup>1</sup>Department of Physiology, University of California, San Francisco, California 94158-2517, USA. <sup>2</sup>Centro de Ecología, Laboratorio de Biología de Organismos, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas 1020-A, Venezuela. <sup>3</sup>Department of Embryology, Carnegie Institution, Baltimore, Maryland 21218, USA. <sup>4</sup>Centro de Biofísica y Bioquímica, Laboratorio de Fisiología Celular, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas 1020-A, Venezuela. <sup>5</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158-2517, USA. <sup>6</sup>California Institute for Quantitative Biosciences, University of California, San Francisco, California 94158-2517, USA. <sup>7</sup>Howard Hughes Medical Institute, University of California, San Francisco, California 94158-2517, USA.

\*These authors contributed equally to this work.





**Figure 2 | Sequence and distribution of vampire bat TRPV1.** **a**, Deduced protein sequences for TRPV1 short (TRPV1-S) and long (TRPV1-L) isoforms from vampire bat sensory ganglia. **b**, Abundance of TRPV1 isoform transcripts in vampire and fruit bat sensory ganglia as determined by transcriptome profiling and/or direct sequencing of RT-PCR products ( $\geq 86$  clones) amplified from sensory ganglia mRNA. **c**, *In situ* hybridization, showing expression of TRPV1 and TRPA1 transcripts in histological sections from vampire bat TG or DRG (scale bar, 50  $\mu$ m). **d**, Prevalence of TRPV1 or TRPA1 mRNA expression within vampire bat TG or DRG (mean  $\pm$  s.e.m.). Data derived from  $n = 554$  neurons, 14 sections for V1-TG; 783 neurons, 17 sections for A1-TG; 1,455 neurons, 15 sections for V1-DRG; and 1,030 neurons, 10 sections for A1-DRG.

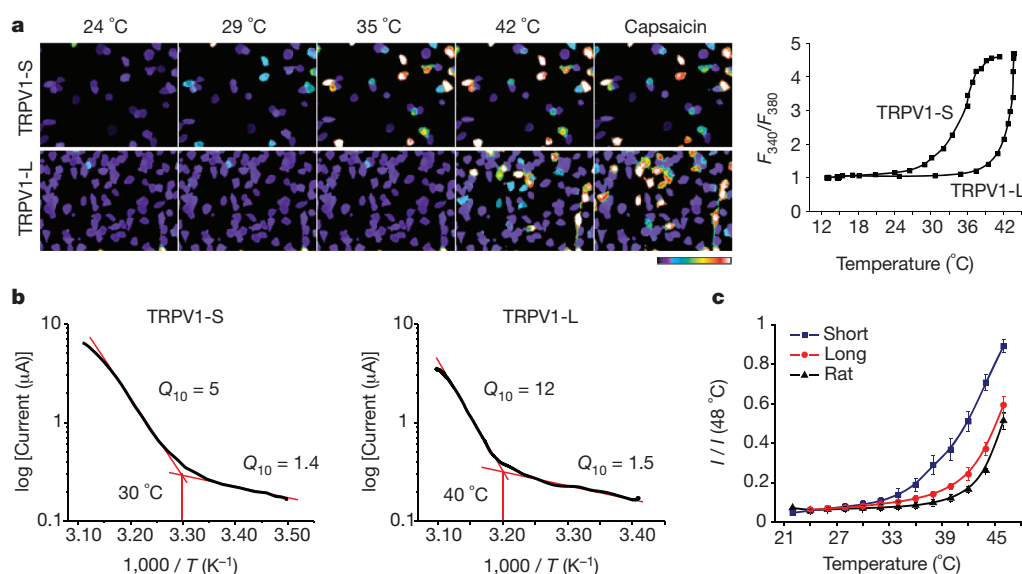
(Fig. 2b), and represented  $\leq 1.8\%$  of TRPV1 transcripts from TG of three other fruit, nectar or insect feeding bat species (*Uroderma bilobatum*, *Sturnira lilium* and *Anoura cultrata*). Taken together, these observations support the notion that TRPV1-S contributes to the

specialized function of vampire bat TG, most notably infrared sensation.

Discrete cellular expression of TRPV1 in vampire bat sensory ganglia was confirmed by *in situ* hybridization. Consistent with deep sequencing analysis, TRPV1 transcripts were found in similar percentages of TG and DRG neurons ( $31.1 \pm 1.5\%$  and  $49.2 \pm 1.5\%$ , respectively) (Fig. 2c, d). Specificity of this pattern was confirmed by examining the distribution of TRPA1 transcripts, which were present in a larger percentage of TG and DRG neurons ( $43.6 \pm 1.7\%$  and  $75.4 \pm 1.0\%$ , respectively) compared to TRPV1 (Fig. 2c, d and Supplementary Fig. 3).

If TRPV1-S contributes to infrared detection in vampire bats, then this isoform should have appropriate temperature sensitivity. We expressed and characterized TRPV1-S and TRPV1-L in HEK293 cells or *Xenopus* oocytes using calcium imaging or electrophysiological assays, respectively. Indeed, we observed a marked difference in temperature sensitivity, such that the TRPV1-S isoform was activated at a substantially lower threshold compared to TRPV1-L ( $30.5 \pm 0.7^\circ\text{C}$  versus  $39.6 \pm 0.4^\circ\text{C}$  in HEK293 cells, and  $31.2 \pm 1.5^\circ\text{C}$  versus  $40.2 \pm 0.7^\circ\text{C}$  in oocytes) (Fig. 3a–c and Supplementary Figs 4, 5). Similar thresholds were observed for fruit bat TRPV1 isoforms (Supplementary Figs 4, 6). This  $\sim 10^\circ\text{C}$  threshold differential between TRPV1-S and TRPV1-L, together with unique expression of the short isoform in vampire bat TG, is consistent with a role for TRPV1-S in infrared detection. In contrast, heterologously expressed TRPA1 channels from either bat species were heat insensitive (Supplementary Fig. 7).

Our histological analysis does not allow us to discriminate between TRPV1 isoforms. We therefore considered whether their co-expression would produce channels having lowered thermal thresholds compared to TRPV1-L alone. Injection of oocytes with equal amounts of TRPV1-S and TRPV1-L RNAs produced an intermediate temperature response curve with an activation threshold of  $33.9 \pm 1.2^\circ\text{C}$  (Supplementary Fig. 5), rather than biphasic thresholds, suggesting that short and long isoforms can produce functional heterotetrameric complexes. Oocytes expressing mostly one isoform (10S:1L or 1S:10L) showed a thermal threshold defined by the predominant species (Supplementary Fig. 5). Thus, if individual TG neurons express equivalent levels of short and long isoforms, then they will have lower thermal thresholds compared to cells predominately expressing the long form.

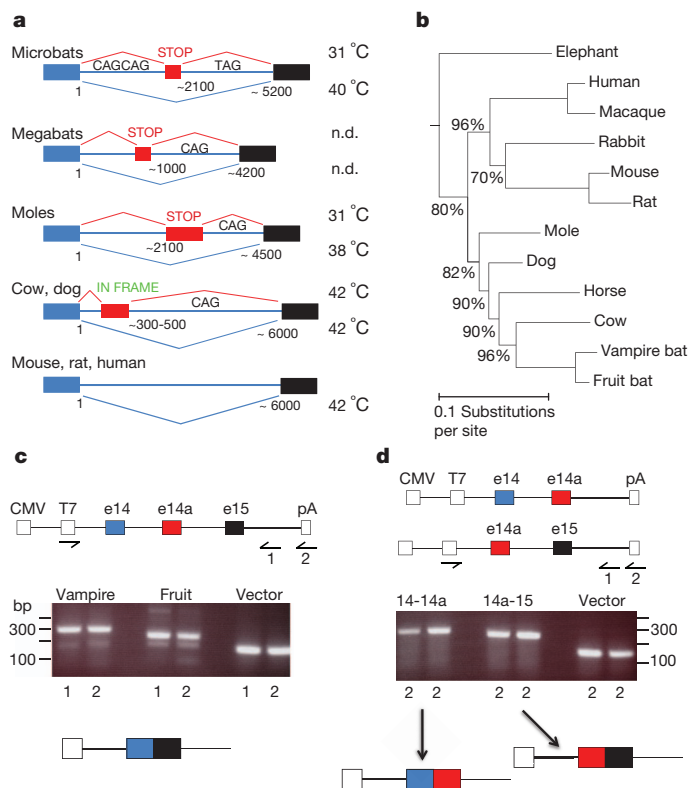


**Figure 3 | Functional analyses of vampire bat TRPV1 isoforms.** **a**, HEK293 cells expressing vampire bat TRPV1 isoforms were analysed for heat or capsaicin (10  $\mu$ M)-evoked responses using calcium imaging; colour bar indicates relative change in fluorescence ratio, with purple and white denoting lowest and highest cytoplasmic calcium, respectively ( $n \geq 141$  cells per channel). Average temperature-response profiles for capsaicin-sensitive cells

are shown at right. **b**, Arrhenius plots show thermal thresholds and  $Q_{10}$  values for baseline and evoked responses for TRPV1-S and TRPV1-L (+80 mV,  $n = 8$ ). **c**, Relative heat response profiles of vampire bat TRPV1 isoforms compared with rat TRPV1, as measured electrophysiologically in oocytes (response at each temperature was normalized to maximal response at  $48^\circ\text{C}$ ;  $V_H = -80$  mV,  $n \geq 8$ ). Data show mean  $\pm$  s.d.

The identification of C-terminal TRPV1 splice variants in vampire bats is somewhat surprising because, to our knowledge, such isoforms have not been previously reported for other species. Alignment between vampire bat TRPV1-S and -L cDNAs highlighted a 23-base-pair insertion in the former (Supplementary Fig. 8a) containing a stop codon that accounts for production of the short isoform. We used sequences flanking this insert to amplify and characterize the organization of the vampire bat *Trpv1* locus in this vicinity. The resulting 5.5-kilobase (kb) fragment contained a tiny 23-base-pair exon (14a) flanked by two introns (2.1 and 3.1 kb), each marked by canonical –GT/AG– donor-acceptor sites required for U2-dependent splicing, as well as obligate polypyrimidine tracts preceding potential splicing sites<sup>16,17</sup>, allowing for formation of two splice variants that incorporate or bypass exon 14a to produce TRPV1-S or TRPV1-L, respectively (Supplementary Fig. 9).

To determine whether exon 14a-based splicing is a hallmark of other Laurasiatheria orders, we carried out transcriptome and/or genomic analysis from Cetartiodactyla (cow, pig), Carnivora (dog) and Lipotyphla (coast mole), as well as several additional bat species representing both Chiroptera suborders, including microbats and megabats<sup>18–20</sup>. Two types of TRPV1 splicing events can occur (Fig. 4a): in microbats, megabats and moles, exon 14a contains a premature stop



**Figure 4 | Genomic organization of mammalian *Trpv1* locus.** **a**, Schematic of *Trpv1* gene locus spanning putative exons 14 and 15 in animals of Laurasiatheria, Rodentia and Primate groups. Splicing events between exons 14 (blue) and 15 (black) are shown, including those involving exon 14a (red). Thermal activation thresholds for resulting channel isoforms are shown at right (n.d., not determined). Lengths of intronic region and relative positions of exon 14a are indicated. **b**, Consensus phylogenetic tree from Bayesian estimation, with clade credibility values shown for branches with <100% support. **c**, **d**, Structures of mini-genes (top) used for *in vivo* splicing assay in HEK293 cells, showing location of CMV promoter and polyadenylation site (pA), defining start and end of transcribed unit. Reverse transcripts were generated with primers annealing to 3' vector sequence (1) or oligo-dT (2), as indicated, and products amplified with T7 and primer (1) pair. Reaction products were resolved on agarose gel (middle) and major bands collected for characterization of splicing products (bottom).

codon accounting for production of a low-threshold TRPV1-S isoform (Supplementary Figs 4–6 and 10). In cows, pigs and dogs, exon 14a produces an in-frame insertion generating an extra-large isoform (TRPV1-XL). Thus, although members of the Laurasiatheria superorder exploit the same intronic region for TRPV1 modification, different suborders show distinct sequences and positions of exon 14a (relative to exons 14 and 15), indicative of independent evolutionary events leading to modification of this intronic region. Interestingly, the architecture of splice sites in microbats differs from that of other Laurasiatheria members, in that the pyrimidine-rich tract is followed by a strong tandem splice site (CAGCAG), and a weak exon 15 acceptor site (TAG in microbats versus CAG in most other species) (Fig. 4a). Consistent with our failure to observe TRPV1-S-like isoforms in mouse sensory ganglia (not shown), neither rodent nor human *TRPV1* gene contains an exon 14a equivalent, limiting splicing to a single isoform.

Although cows and moles have the potential to produce TRPV1 splice variants, we found that <6% of TRPV1 transcripts corresponded to TRPV1-XL or TRPV1-S in TG of cow or mole, respectively, indicating that this form of post-transcriptional regulation is not physiologically relevant in these animals. In cows, this is further underscored by the fact that TRPV1-L and TRPV1-XL isoforms are indistinguishable in regard to thermal response profiles (thresholds of  $43.0 \pm 0.8^\circ\text{C}$  and  $42.7 \pm 0.4^\circ\text{C}$ , respectively) (Supplementary Figs 4 and 11). Our analysis of *Trpv1* gene organization in different orders (including Chiroptera, Cetartiodactyla, Carnivora, Lipotyphla, Afrotheria and Rodentia) (Fig. 4b and Supplementary Figs 9 and 12) is consistent with an independent molecular phylogeny of mammals<sup>21,22</sup> and supports the conclusion that bats are more closely related to cows, moles and dogs<sup>19,20</sup>, rather than to rodents (as initially posited on the basis of anatomical and morphological criteria<sup>23</sup>).

*Trpv1* genomic sequences for vampire and other microbats in the vicinity of exon 14–14a–15 splice junctions are highly conserved (>90% identity). We therefore asked whether differential splicing was observed when mini-genes containing these genomic regions were introduced into HEK293 cells. Direct splicing of exon 14 to 15, with exclusion of exon 14a, was observed in all cases (Fig. 4c). In mini-gene constructs where direct splicing between exon 14 and 15 was not possible, splicing to or from exon 14a was readily observed (Fig. 4d), indicating that exon 14a is competent to engage in splicing, but disfavoured by exon 14–15 competition. These results show that vampire and fruit bat genes exhibit the same default pattern of exon 14 to 15 splicing in this non-neural context, irrespective of small differences in gene structure. Indeed, this ‘constitutive’ splicing pattern predominates in vampire bat DRG, fruit bat TG and DRG, as well as cow and mole TG, suggesting that efficient splicing to exon 14a in vampire bat TG requires a specialized environment. Furthermore, we have used this assay to identify a putative exon 14a from megabat species, such as *Pteropus vampyrus*, where transcriptome data are not available (Fig. 4a and Supplementary Fig. 13).

Our findings suggest that variation within the TRPV1 C-terminus represents a genetic mechanism for tuning the thermal response profile of the channel in a species- or tissue-specific manner. We therefore asked whether more primitive and evolutionarily distant species use a similar strategy for thermosensory adaptation. Indeed, we found that zebrafish TRPV1 is activated with a threshold of  $32.9 \pm 1.2^\circ\text{C}$  (Supplementary Fig. 14a–d), consistent with physiological adaptation of the animal to its environment of  $25\text{--}33^\circ\text{C}$  (ref. 24). When compared to rat TRPV1, the zebrafish channel has a gap of 12 amino acids within the C terminus corresponding precisely to the location of the splice junction in the vampire bat channel (Supplementary Fig. 14e). This gap results from a polymorphism within exon 15 of the zebrafish *trpv1* gene, and thus zebrafish and vampire bats use different mechanisms to generate TRPV1 C-terminal variants.

RNA splicing extends the coding potential of the genome and enhances functionality of the proteome<sup>25</sup>. The vampire bat uses this strategy to produce physiologically distinct channels, thereby generating

a hypersensitive detector (TRPV1-S) within a specific thermosensory organ, without sacrificing somatic thermo-sensation and/or thermo-nociception. Moreover, our results demonstrate that sequence variation within a specific region of the cytoplasmic C-terminus accounts for differential thermal response profiles of vampire bat TRPV1 isoforms. Indeed, we previously proposed that this region of the channel interacts with membrane phospholipids to modulate sensitivity to thermal and chemical stimuli<sup>26</sup>.

In addition to illuminating mechanisms of thermosensation and sensory adaptation, the analysis of TRP channel gene structure provides a physiologically relevant marker for assessing phylogenetic relationships. Our findings support recent molecular classification in which bats are grouped together with horses, dogs, cows, moles and dolphins (Laurasiatheria superorder), rather than with humans, monkeys, flying lemur, mouse, rat and rabbits (Euarchontoglires superorder) as originally proposed on the basis of anatomical criteria.

## METHODS SUMMARY

Bat tissue was fixed with paraformaldehyde for *in situ* hybridization histochemistry or homogenized in TRIzol for nucleic acid analysis. cDNA libraries were sequenced on Illumina Genome Analyzer II. Cloned channels were transiently expressed in HEK293 cells and subjected to calcium imaging using Fura-2/AM ratiometric dye. *Xenopus* oocytes were cultured, injected with 5 ng of cRNA, and analysed 2–5 d post-injection by two-electrode voltage clamp as described<sup>27</sup>. Thermal stimulation was applied with a custom-made Peltier device (Reid-Dan Electronics). Temperature thresholds represent the point of intersection between linear fits to baseline and the steepest component of Arrhenius profile, as described<sup>28</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 January; accepted 3 June 2011.

- Kurten, L. & Schmidt, U. Thermoperception in the common vampire bat (*Desmodus rotundus*). *J. Comp. Physiol.* **146**, 223–228 (1982).
- Schutt, B. *Dark Banquet: Blood and the Curious Lives of Blood-Feeding Creatures* (Three Rivers Press, 2008).
- Tellgren-Roth, A. *et al.* Keeping the blood flowing — plasminogen activator genes and feeding behavior in vampire bats. *Naturwissenschaften* **96**, 39–47 (2009).
- Campbell, A. L., Naik, R. R., Sowards, L. & Stone, M. O. Biological infrared imaging and sensing. *Micron* **33**, 211–225 (2002).
- Bullock, T. H. & Cowles, R. B. Physiology of an infrared receptor: the facial pit of pit vipers. *Science* **115**, 541–543 (1952).
- Gracheva, E. O. *et al.* Molecular basis of infrared detection by snakes. *Nature* **464**, 1006–1011 (2010).
- Kurten, L., Schmidt, U. & Schafer, K. Warm and cold receptors in the nose of the vampire bat *Desmodus rotundus*. *Naturwissenschaften* **71**, 327–328 (1984).
- Molenaar, G. J. The sensory trigeminal system of a snake in the possession of infrared receptors. II. The central projections of the trigeminal nerve. *J. Comp. Neurol.* **179**, 137–151 (1978).
- Bakken, G. S. & Krochmal, A. R. The imaging properties and sensitivity of the facial pits of pitvipers as determined by optical and heat-transfer analysis. *J. Exp. Biol.* **210**, 2801–2810 (2007).
- Safer, A. B. & Grace, M. S. Infrared imaging in vipers: differential responses of crotaline and viperine snakes to paired thermal targets. *Behav. Brain Res.* **154**, 55–61 (2004).
- Kishida, R., Goris, R. C., Terashima, S. & Dubbeldam, J. L. A suspected infrared-recipient nucleus in the brainstem of the vampire bat, *Desmodus rotundus*. *Brain Res.* **322**, 351–355 (1984).
- Schafer, K., Braun, H. A. & Kurten, L. Analysis of cold and warm receptor activity in vampire bats and mice. *Pflügers Arch.* **412**, 188–194 (1988).
- Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
- Jordt, S. E., McKemy, D. D. & Julius, D. Lessons from peppers and peppermint: the molecular logic of thermosensation. *Curr. Opin. Neurobiol.* **13**, 487–492 (2003).
- Ramsey, I. S., Delling, M. & Clapham, D. E. An introduction to TRP channels. *Annu. Rev. Physiol.* **68**, 619–647 (2006).
- Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
- Will, C. L. & Luhrmann, R. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.* **386**, 713–724 (2005).
- Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–618 (2001).
- Murphy, W. J. *et al.* Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351 (2001).
- Nishihara, H., Hasegawa, M. & Okada, N. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl Acad. Sci. USA* **103**, 9929–9934 (2006).
- Asher, R. J., Bennett, N. & Lehmann, T. The new framework for understanding placental mammal evolution. *Bioessays* **31**, 853–864 (2009).
- Prasad, A. B., Allard, M. W. & Green, E. D. Confirming the phylogeny of mammals by use of large comparative sequence data sets. *Mol. Biol. Evol.* **25**, 1795–1808 (2008).
- Pettigrew, J. D. *et al.* Phylogenetic relations between microbats, megabats and primates (Mammalia: Chiroptera and Primates). *Phil. Trans. R. Soc. Lond. B* **325**, 489–559 (1989).
- Spence, R., Gerlach, G., Lawrence, C. & Smith, C. The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol. Rev. Camb. Phil. Soc.* **83**, 13–34 (2008).
- Grabowski, P. J. & Black, D. L. Alternative RNA splicing in the nervous system. *Prog. Neurobiol.* **65**, 289–308 (2001).
- Prescott, E. D. & Julius, D. A modular PIP2 binding site as a determinant of capsaicin receptor sensitivity. *Science* **300**, 1284–1288 (2003).
- Chuang, H. H., Neuhauser, W. M. & Julius, D. The super-cooling agent icilin reveals a mechanism of coincidence detection by a temperature-sensitive TRP channel. *Neuron* **43**, 859–869 (2004).
- DeCoursey, T. E. & Cherny, V. V. Temperature dependence of voltage-gated H<sup>+</sup> currents in human neutrophils, rat alveolar epithelial cells, and mammalian phagocytes. *J. Gen. Physiol.* **112**, 503–522 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank Y. Kelly and J. Poblete for technical assistance, C. Sehnert for help with bovine tissue collection, M. Suzawa and H. Ingraham for providing zebrafish mRNA, A. Walsh for providing megabat blood samples, the Centro Técnico de Producción Socialista Florentino for providing access to Hato Piñero (Cojedes, Venezuela) for animal collection and J. Nassar for providing access to laboratory material required for specimen collection. This work was supported by a Ruth L. Kirschstein National Research Service Award (GM080853; N.T.I.), a Pathway to Independence Fellowship from the UCSF CVRI (E.O.G.), the Howard Hughes Medical Institute (J.S.W.), and grants from NIH, including P01 AG010770 (J.S.W.) and NS047723 and NS055299 (D.J.).

**Author Contributions** E.O.G., J.F.C.-M. and N.T.I. designed and performed experiments and analysed data. N.T.I. and J.S.W. developed analytical tools and analysed data. J.A.G.-C., C.I.A. and C.M. collected bat species and obtained tissues for analysis. E.O.G., J.F.C.-M. and D.J. wrote the manuscript with discussion and contributions from all authors. J.S.W. and D.J. provided advice and guidance throughout.

**Author Information** Deep sequencing data are archived under GEO accession number GSE28243. GenBank accession numbers are JN006855 (*D. rotundus* TRPV1-S), JN006856 (*D. rotundus* TRPV1-L), JN006857 (*D. rotundus* TRPA1), JN006858 (*C. brevicauda* TRPA1), JN006859 (*C. brevicauda* TRPV1-L), JN006860 (*C. brevicauda* TRPV1-S), JN006861 (*Scapanus orarius* TRPV1-L), JN006862 (*S. orarius* TRPV1-S), JN006863 (*Pteropus rodricensis* intron), JN006864 (*D. rotundus* intron), JN006865 (*C. brevicauda* intron), JN006866 (*P. vampyrus* intron), JN006867 (*Rousettus aegyptiacus* intron) and JN006868 (*S. orarius* intron). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.J. ([david.julius@ucsf.edu](mailto:david.julius@ucsf.edu)) or to N.T.I. ([ingolia@ciwemb.edu](mailto:ingolia@ciwemb.edu)) for information concerning bioinformatics.



## METHODS

**Tissue collection.** Vampire bats (*D. rotundus*) and fruit/nectar/insect eating bats (*C. brevicauda*, *A. cultrata*, *U. bilobatum* and *S. lilium*) were collected in pasture forests grazed by cattle at the Hato Piñero ranch, property of the Centro Técnico de Producción Socialista Florentino, located near El Baúl, Cojedes State, Venezuela. Genetic resource and animal sampling permits were obtained from the Ministerio del Ambiente (MINAM). Global coordinates for animal collection are provided in Supplementary Fig. 15.

Bovine trigeminal ganglia were collected from 12-day old calves for RNA extraction under supervision of C. Sehnert. Blood samples from megabats were generously provided by the Lube Bat Conservatory. A summary of samples from all species examined is provided in Supplementary Fig. 15.

**Deep sequencing and analysis.** Sequencing libraries were prepared from poly (A)+ RNA using the Illumina mRNA-Seq Sample Prep Kit according to the manufacturer's instructions. Libraries were then sequenced on the Illumina Genome Analyzer II using two 36-cycle sequencing kits to read 80 nucleotides of sequence from a single end of each insert, by standard protocols. Between 7.6 and 26.2 million inserts were sequenced for each sample.

Sequences were aligned to the cow RefSeq protein database (downloaded from NCBI ftp://ftp.ncbi.nih.gov/genomes/Bos\_taurus/protein/protein.fa.gz on 25 August 2009) using the blastx tool from NCBI blast (version 2.2.21), which aligns a six-frame translation of each query against a protein database. The alignment was performed with a word size of four amino acids and a window size of five; a maximum *E* value of  $1 \times 10^{-5}$  was required. For each read that aligned to the cow proteome, a set of optimal hits was collected based on alignments whose bit score was within 0.2 of the highest bit score reported for that sequencing read. Each RefSeq alignment for a given sequencing read was converted to an Entrez Gene identifier and redundant alignments for a single read, which correspond to alignments against different isoforms of the same protein, were collapsed to a single count. The number of optimally aligning reads was then counted for each gene; in some cases a single read counted towards several genes.

Splice variants were quantified by aligning sequencing reads to the nucleotide sequence of the alternative TRPV1 isoforms using Bowtie v0.11.3. The alignments were manually inspected to ensure that they supported the splice junction with at least 24 nucleotides aligning on one side of the TRPV1 transcript and at least two nucleotides unambiguously distinguishing the isoform.

The TRPV1 protein sequences were taken from this study, Ensembl (finch, macaque, horse and elephant) and RefSeq (all others), and aligned using t\_coffee (ref. 29) in 'accurate' mode with the blosum80mt distance matrix. The protein multiple sequence alignment was used as input to MrBayes<sup>30</sup> and the phylogeny was estimated using gamma-distributed substitution rates under the Blosum62 distance matrix combined with a fraction of invariant sites. The phylogeny distribution was estimated across 10 runs, each of 6,000,000 MCMC cycles, with samples every 100,000 cycles following a 1,000,000 cycle burn-in. The chicken and finch sequences were used to root the mammalian tree and are not shown.

**Cloning.** Functional cDNAs were amplified from first strand cDNA generated by reverse transcription using following primers. Vampire bat *Trpv1* (forward: 5'-GCAAGGATGAAGAAACGGG-3'; reverse: 5'-CCTCACTTTTCCCCTAAA GC-3'). Vampire bat *Trpa1* (forward: 5'-GGCCTGCGTAACATCAGAAGC-3'; reverse: 5'-TACTGCTAAGGTCATGCACTTTTGG-3'). *C. brevicauda* *Trpv1* (forward: 5'-CTTGGGCAAAGATGAAGAAACG-3'; reverse: 5'-GTGTGCTAG TCTAAGGCCA-3'). *C. brevicauda* *Trpa1* (forward: 5'-GCATCCAGGGTAGG ATCCAT-3'; reverse: 5'-CAAGAAACGTGTGTGTGGGA-3'). Zebrafish *trpv1* (forward: 5'-CCTCAAGCCAAGTTACTCAC-3'; reverse: 5'-TCGAAGGACAC CTTGTAGAC-3'). Cow *Trpv1* (forward: 5'-ATGAAGAAATGGGGAGCT CAG-3'; reverse: 5'-TCACCTTCTCCCTAAAGCCAC-3'). Coast mole *Trpv1* (forward: 5'-GGAGCTGGCAAGGATGAAG-3'; reverse: 5'-ATCCTAAGGCC CAACAGAGT-3'). Genomic DNA was extracted from vampire bat (*Desmodus rotundus*) and fruit bat (*C. brevicauda*, *Sturnira bilobatum*) brain, and coast mole

TG, using TRIzol Reagent (Invitrogen); genomic DNA of megabats was extracted from blood using QIAmp DNA Blood mini kit (Qiagen). The genomic region between exon 14 and 15 was amplified using the following primers. Vampire bat, megabats (forward: 5'-CAGGCAATTGTGAAGGCATC-3'; reverse: 5'-CCA GGGCAAAGTCTTCCAG-3'). *C. brevicauda*, *Sturnira bilobatum* (forward: 5'-CGGGCAATTGTGAAGGCAT-3'; reverse: 5'-CCAGGGCAAAGTCTT CCAG-3'). Coast mole (forward: 5'-CTTCTCCCTGCGGTCAAG-3'; reverse: 5'-CTCGAGTGCTTGCATCTCTTAA-3').

**Oocyte electrophysiology.** Surgically extracted oocytes from *Xenopus laevis* (Nasco) were cultured, injected with 5 ng of RNA, and analysed 4–5 d post-injection by two-electrode voltage-clamp using a Geneclamp 500 amplifier (Axon Instruments, Inc.) as described<sup>27</sup>. Microelectrodes were pulled from borosilicate glass capillary tubes to obtain resistances of 0.3–0.7 MΩ. Bath solution contained 10 mM HEPES, 120 mM NaCl, 2 mM KCl, 1 mM EGTA and 2 mM MgCl<sub>2</sub> buffered to a final pH of 7.4 with NaOH. Data were analysed using pCLAMP10 software.

**Determination of thermal threshold and temperature coefficients (Q<sub>10</sub>).** Temperature thresholds represent the point of intersection between linear fits to baseline and the steepest component of the Arrhenius profile. Values are derived from averages of individual curves; *n* ≥ 8. Arrhenius curve were obtained by plotting the current (+80 mV) on a log scale against the reciprocal of the absolute temperature. Q<sub>10</sub> was used to characterize the temperature dependence of the ionic current as calculated using the following equation:

$$Q_{10} = \left[ \frac{R_2}{R_1} \right]^{10/(T_2 - T_1)}$$

where *R*<sub>2</sub> is the current at the higher temperature (*T*<sub>2</sub>) and *R*<sub>1</sub> is the current at the lower temperature (*T*<sub>1</sub>) (ref. 28).

**Calcium imaging.** All tested channels were transiently expressed in HEK293 cells with the use of Lipofectamin 2000 (Invitrogen). Calcium imaging of HEK293 cells using Fura-2/AM was performed on coverslips coated with Matrigel (BD). Fluorescent images were acquired with Metafluor software (Molecular Devices) and analysed using Graph Pad Prism 4.

**In situ hybridization histochemistry.** TG and DRG tissue were dissected and fixed in 4% PFA in PBS for 2 d. Cryostat sections (12–15 μm thick) were processed and probed with a digoxigenin-labelled cRNA. Probes were generated by T7/T3 *in vitro* transcription reactions using full length *Trpv1* cDNA and a 3-kb fragment of *Trpa1* cDNA. Signal was developed with alkaline phosphatase-conjugated anti-digoxigenin Fab fragments, according to the manufacturer's instructions (Roche).

**In vivo splicing assay.** Mini-gene constructs. Genomic region containing vampire bat or *C. brevicauda* splice junction between exon 14–15, exon 14–14a and exon14a–15 was topo-cloned, then subcloned into pcDNA3(+) vector (Invitrogen) using BamHI/XhoI restriction sites. Transfections. HEK293 cells were cultured in DMEM medium supplemented with 10% fetal bovine serum in 6-well plates. At 50% confluence, cells were transiently transfected with 1 μg of mini-gene plasmid or empty pcDNA3 (+) vector using the Lipofectamin method. Cells were harvested 48h after transfection using TRIzol reagent (Invitrogen). cDNA synthesis and RT-PCR. RNA was extracted using TRIzol reagent. RNA (10 μg) was used as template for cDNA synthesis using SuperScriptIII first strand synthesis system (Invitrogen). Synthesis was primed with oligo (dT) or by use of primer annealing to 3'-untranslated sequences in pcDNA3 (CAGTCGAGGCT GATCAGCGAGCT). 5% of cDNA was used in 36-cycle PCR using forward: 5'-TAATACGACTCACTATAGGG-3'; reverse: 5'-CAGTCGAGGCTGATCAG CGAGCT-3' primers. Reaction products were resolved by electrophoresis on 1.5% agarose gel. PCR products were isolated from the gel, cloned into Strataclone vector and sequenced.

29. Notredame, C., Higgins, D. G. & Heringa, J. J., *Mol. Biol.* **302**, 205–217 (2000).

30. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

# Photoentrainment and pupillary light reflex are mediated by distinct populations of ipRGCs

S.-K. Chen<sup>1</sup>, T. C. Badea<sup>2</sup> & S. Hattar<sup>1,3</sup>

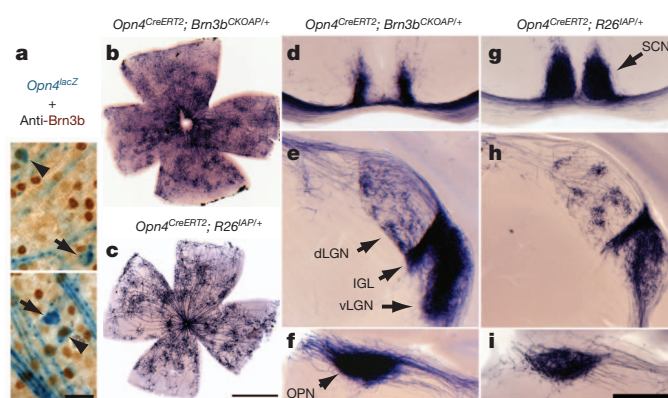
**Intrinsically photosensitive retinal ganglion cells (ipRGCs) express the photopigment melanopsin and regulate a wide array of light-dependent physiological processes<sup>1–11</sup>. Genetic ablation of ipRGCs eliminates circadian photoentrainment and severely disrupts the pupillary light reflex (PLR)<sup>12,13</sup>. Here we show that ipRGCs consist of distinct subpopulations that differentially express the *Brn3b* transcription factor, and can be functionally distinguished. *Brn3b*-negative M1 ipRGCs innervate the suprachiasmatic nucleus (SCN) of the hypothalamus, whereas *Brn3b*-positive ipRGCs innervate all other known brain targets, including the olivary pretectal nucleus. Consistent with these innervation patterns, selective ablation of *Brn3b*-positive ipRGCs severely disrupts the PLR, but does not impair circadian photoentrainment. Thus, we find that molecularly distinct subpopulations of M1 ipRGCs, which are morphologically and electrophysiologically similar, innervate different brain regions to execute specific light-induced functions.**

In addition to rod and cone photoreceptors, the retina contains a small subset of ipRGCs that express the photopigment melanopsin<sup>1,9</sup>. ipRGCs project to the suprachiasmatic nucleus (SCN) and the olivary pretectal nucleus (OPN), regions in the brain that control circadian rhythms and the pupillary light reflex (PLR), respectively. In the absence of the melanopsin protein (*Opn4*), ipRGCs lose their intrinsic photosensitivity<sup>7</sup>, but still innervate the correct brain regions<sup>7</sup> and convey rod/cone input<sup>7,14,15</sup> to drive non-image-forming visual functions<sup>7,16</sup>. Recent studies have shown that ipRGCs are not uniform and can be further subdivided into distinct subtypes based on their morphology, electrophysiology and discrete brain targets<sup>2,17</sup>. M1 ipRGCs can be readily distinguished from other ipRGC subtypes (herein referred to as non-M1 ipRGCs) because they are the only subtype with exclusive dendritic stratification in the OFF sublamina of the inner plexiform layer (IPL) in the retina<sup>18,19</sup>. The prevailing view is that M1 ipRGCs are a homogeneous population that send collateral axonal branches to two relay nuclei, the SCN and OPN, to drive circadian photoentrainment and PLR<sup>20</sup>. Genetic ablation of ipRGCs by diphtheria toxin (in *Opn4*<sup>ΔDTA</sup> mice) eliminates circadian photoentrainment and disrupts PLR<sup>12</sup>. Here we surprisingly found that M1 ipRGCs are not a uniform population, but consist of functionally distinct subpopulations defined by their expression of the POU domain transcription factor, *Brn3b*.

Previously, we showed that *Brn3b* (also known as *Pou4f2*) mutant mice, which lack 80% of RGCs, have pronounced deficits in PLR, but are still capable of weak photoentrainment<sup>21</sup>. These findings raise the possibility that the remaining *Brn3b*-negative M1 ipRGCs selectively mediate photoentrainment. To determine the extent of *Brn3b* expression in the M1 ipRGC population, we performed anti-*Brn3b* immunohistochemistry on retinas from *Opn4*<sup>lacZ</sup> mice, together with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) staining that labels only M1 ipRGCs<sup>21</sup>. A fraction of β-galactosidase-positive RGCs stained for *Brn3b* in the adult retina (Fig. 1a).

To determine the projections of *Brn3b*-positive ipRGCs, we mated mice in which an inducible Cre recombinase is driven by the melanopsin

promoter (*Opn4*<sup>CreERT2/+</sup>) to mice having either a ubiquitous Cre-dependent alkaline phosphatase (AP) reporter (*Rosa26-IAP*)<sup>22</sup> or a conditional *Brn3b* knock-in (*Brn3b*<sup>CKOAP/+</sup>)<sup>21</sup> in which Cre recombination causes the alkaline phosphatase coding region to be expressed by the *Brn3b* promoter (Supplementary Fig. 1). Tamoxifen injections in *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> animals result in labelling of M1 and non-M1 ipRGCs by alkaline phosphatase histochemistry (Fig. 1c), but only of *Brn3b*-expressing ipRGCs in *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> animals (Fig. 1b and Supplementary Table 1). Alkaline phosphatase labelling of *Brn3b*-positive ipRGCs allowed us to analyse the dendritic arborizations and central projections of these cells, independently of *Brn3b*-negative ipRGCs (Fig. 1b, d–f). Many *Brn3b*-positive ipRGCs had dendrites arborizing in the ON sublamina of the IPL similar to previous observations for non-M1 ipRGCs<sup>2,18,20,23</sup>. This indicates that *Brn3b* expression is not just restricted to M1 ipRGCs, but is also expressed in non-M1 ipRGCs (Fig. 1b). Comparing the labelling of *Brn3b*-positive M1 and non-M1 ipRGCs with all ipRGC subtypes, we find that most brain targets of ipRGCs show similar patterns of innervation (Fig. 1d–i). In particular, the OPN is innervated fully in both cases (Fig. 1f, i). However, a notable difference is found in the SCN; in the *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice, the SCN was completely innervated by alkaline-phosphatase-positive ipRGC fibres (Fig. 1g) similarly to previous studies<sup>2</sup>. In contrast, in the *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice,



**Figure 1 | Co-expression of melanopsin and *Brn3b* defines a specific set of ipRGCs.** **a**, Retinal flat mounts from *Opn4*<sup>lacZ</sup> mice stained with anti-*Brn3b* antibody (brown) and X-gal staining (blue) show *Brn3b*-positive (arrowheads; 140 *Brn3b*-positive ipRGCs from 988 lacZ+ cells, *n* = 5) and *Brn3b*-negative (arrows), M1 ipRGCs. **b–i**, Alkaline phosphatase histochemistry of retina (**b** and **c**) and coronal brain sections (**d–i**) from *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice (**b**, **d–f**), or from *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice (**c**, **g–i**). Suprachiasmatic region shows partial innervation in *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice (**d**), compared to full innervation of the SCN in *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice (**g**). Both mouse lines show significant ipRGC projections to the IGL and vLGN, and sparse innervation to the dLGN (**e** and **h**) and intense labelling of the OPN (**f** and **i**). Scale bars are 25 μm (**a**), 1 mm (**b** and **c**), and 400 μm (**d–i**).

<sup>1</sup>Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>2</sup>Retinal Circuit Development & Genetics Unit, N-NRL/NEI/NIH, Bethesda, Maryland 20892, USA. <sup>3</sup>Department of Neuroscience, Johns Hopkins University-School of Medicine, Baltimore, Maryland 21218, USA.

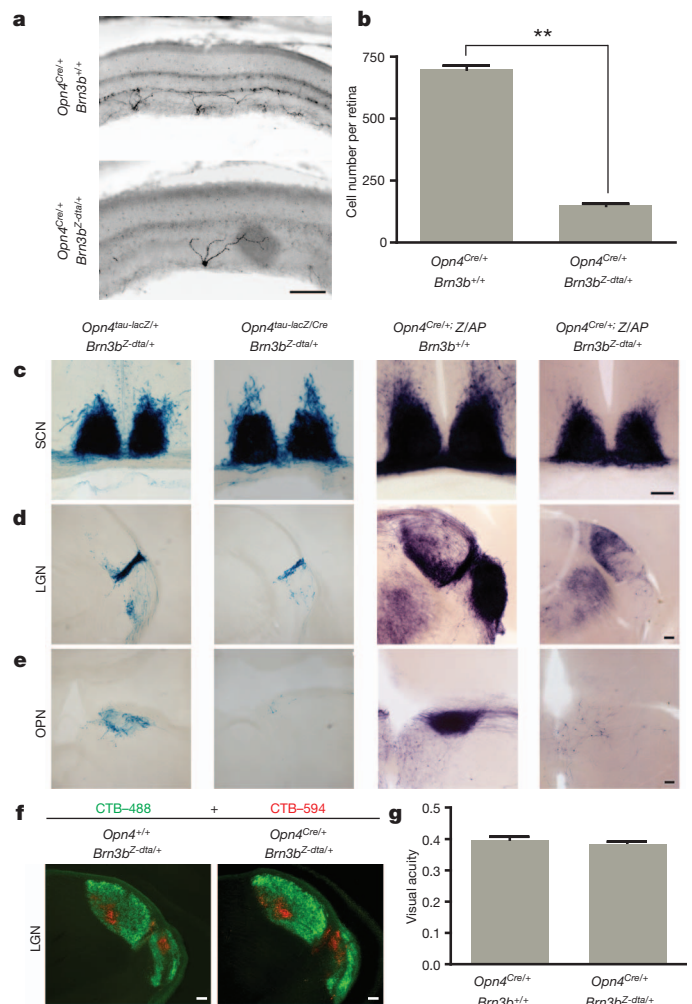


the SCN was sparsely innervated by Brn3b-positive ipRGCs (Fig. 1d), with the medial regions of the SCN completely devoid of innervation (Fig. 1d). We further confirmed that the diminished SCN innervation is not due to the use of the inducible *Opn4<sup>CreERT2</sup>* line, because crossing the *Opn4<sup>Cre</sup>* line<sup>2</sup> with the *Brn3b* knock-in allele (*Opn4<sup>Cre</sup>; Brn3b<sup>CKOAP/+</sup>* animals) also results in reduced SCN innervation (Supplementary Fig. 2). Thus, ipRGCs can be separated into two subpopulations based on their Brn3b expression and connectivity.

To label the central projection of Brn3b-negative ipRGCs and determine the physiological functions of both Brn3b-negative and Brn3b-positive ipRGCs, we specifically eliminated cells that co-express melanopsin and Brn3b by crossing *Opn4<sup>Cre</sup>* and *Brn3b<sup>Z-dta</sup>* lines (Supplementary Table 1). The *Brn3b<sup>Z-dta</sup>* knock-in line<sup>24</sup> expresses diphtheria toxin A subunit (DTA) from the endogenous *Brn3b* gene promoter (Supplementary Fig. 1) only in the presence of Cre<sup>25</sup>. Thus, in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, Brn3b-expressing ipRGCs are ablated, whereas Brn3b-negative ipRGCs and conventional (melanopsin negative) RGCs are left intact (Supplementary Fig. 3). Using melanopsin immunofluorescence that only reveals M1 ipRGCs in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* retinas, we observed less than 200 surviving M1 ipRGCs (Fig. 2a, b). To determine the extent of ablation of all ipRGCs in the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, we generated triple heterozygous *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* mice (Supplementary Table 1). Alkaline phosphatase labelling in the *Opn4<sup>Cre/+</sup>; Z/AP* mice reveals all M1 and non-M1 ipRGCs (~2,000 cells)<sup>2</sup>. Using alkaline phosphatase histochemistry in the triple heterozygous mice, we observed similar numbers of surviving ipRGCs as with melanopsin immunofluorescence (Fig. 2a). These results show that all non-M1 cells are ablated and that the surviving 10% (200 out of 2,000) of total ipRGCs are Brn3b-negative and predominantly belong to the M1 subtype.

To assess the central projections of these surviving M1 Brn3b-negative ipRGCs, we crossed *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice with the either the *Opn4<sup>tau-LacZ</sup>* reporter<sup>19,20</sup> or *Z/AP* reporter<sup>26</sup>. Although only 200 M1 ipRGCs remained in the *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* or *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* mice, we observed that their fibres completely innervated the SCN at levels comparable to those observed in the control groups (Fig. 2c). However, innervation of the intergeniculate leaflet (IGL) was highly attenuated (Fig. 2d) and OPN projections were completely abolished (Fig. 2e). Interestingly, the shell of the OPN showed no fibres in the *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* compared to control mice (Fig. 2e). Given that both the SCN and OPN shell are innervated by M1 ipRGCs<sup>5,20</sup>, differential labelling of these ipRGC targets in *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* mice shows that the M1 subtype of ipRGCs is not a uniform population. To ensure that RGCs that are not intrinsically photosensitive are intact in the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, we used cholera toxin injection in the eye to label all RGC fibres in the brain anterogradely. RGCs innervated the dorsal and ventral lateral geniculate nuclei (dLGN and vLGN) normally in these mice (Fig. 2f). This is further supported by the similar visual acuity measured in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* and wild-type mice (Fig. 2g).

Given that ipRGC projections to the OPN are lost, but SCN projections are largely intact, the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice allow the relative contributions of Brn3b-negative ipRGCs to the pupillary light reflex (PLR) and circadian light responses to be determined. We first measured PLR in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice at two light intensities in the middle of the day (zeitgeber time 8, ZT8). The pupil of wild-type mice is 95.61% constricted under high light intensity and 79.47% under low light intensity (Fig. 3a, c). In contrast, *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice showed a highly attenuated PLR at ZT8 under high and low light intensities (Fig. 3b, c). This phenotype is remarkably similar to the PLR deficits observed in *Opn4<sup>aDTA/aDTA</sup>* homozygous animals<sup>12</sup>. We further investigated the PLR in the middle of the night (ZT20), and found that *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals have no pupillary constriction to high or low light stimulations (Supplementary Fig. 4 and Supplementary Text). The residual PLR response at ZT8

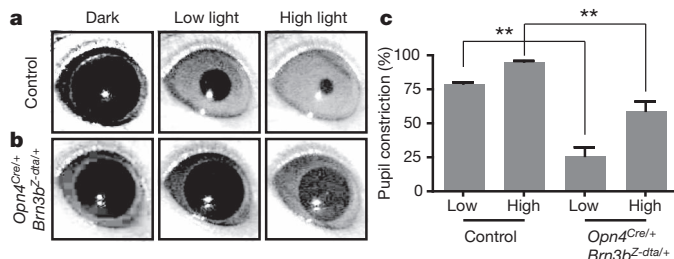


**Figure 2 | Genetic ablation of Brn3b-positive ipRGCs does not impair targeting to the SCN.** **a**, Melanopsin immunofluorescence reveals a reduction in ipRGC numbers in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* retina compared to control (*Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>*). **b**, Quantification of surviving melanopsin-positive cells in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* (149.8 ± 8.65 cells per retina; n = 4) and control (698.8 ± 16.85 cells per retina; n = 4) mice. **c–e**, Coronal brain sections of *Opn4<sup>tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* and *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* (**c–e**, left two panels), and *Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>; Z/AP* and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* (**c–e**, right two panels) mice using X-gal (left two panels) or alkaline phosphatase histochemistry (**c–e**, right two panels). Sections show SCN (**c**), LGN (**d**) and OPN (**e**). **f**, Labelling of all RGCs with Alexa Fluor 594- and 488-conjugated cholera toxin B (CTB-488 and CTB-594, respectively) in the left (red) and right eye (green), respectively, shows normal brain targeting to image forming regions. **g**, Visual acuity was the same between *Opn4<sup>+/+</sup>; Brn3b<sup>Z-dta/+</sup>* (n = 5) and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice (n = 6). Scale bars are 100 μm. Error bars represent s.e.m.

in *Opn4<sup>aDTA/aDTA</sup>* and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice suggests that other melanopsin-negative RGCs contribute to this reflex. One candidate population could be Brn3a-positive RGCs, which project to the OPN<sup>21</sup>.

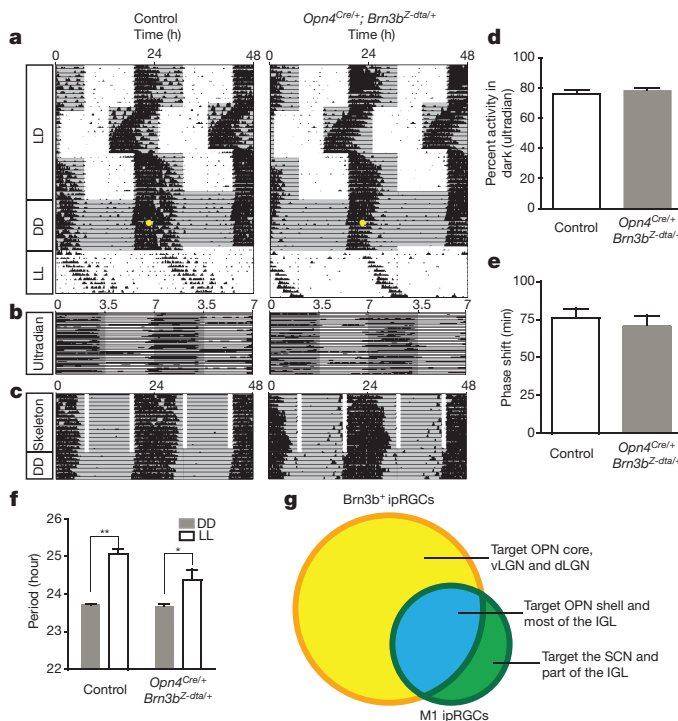
We then asked whether the surviving M1 ipRGCs that innervate the SCN in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice are sufficient to drive circadian photoentrainment. Strikingly, we found that *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice are able to photoentrain as well as controls under normal 24-h light dark cycles or skeleton photoperiods (Fig. 4a, c, Supplementary Figs 5 and 6 and Supplementary Text). In addition, they can readjust to a 'jet lag' light-dark cycle paradigm with advanced and delayed dark onsets (Fig. 4a). We also observed no difference in activity during the dark phase of the ultradian<sup>16</sup> 3.5:3.5 light/dark (LD) cycle (Fig. 4b, d). Moreover, a 15-min light pulse presented early during the active phase





**Figure 3** | *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show severe deficits in the pupillary light reflex (PLR). **a, b**, Representative images of PLR from control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice. Left panels show pupils under dark conditions, middle panels show pupils under low light intensity ( $22 \mu\text{W cm}^{-2}$ ) and right panels show pupils under high light intensity ( $5.66 \text{ mW cm}^{-2}$ ). **c**, Quantification of PLR data from control ( $n = 5$ ) and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* ( $n = 6$ ) animals. \*\* indicates  $P < 0.01$  with one-way ANOVA. Error bars represent s.e.m.

of mice maintained under constant conditions generated similar phase shifts ( $76.2 \pm 6.5$  and  $70.67 \pm 7.3$  min for control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals, respectively; Fig. 4e). Together, these results indicate that Brn3b-negative ipRGCs, comprising only 10% of all identified ipRGC subtypes, are sufficient for circadian photoentrainment.



**Figure 4** | *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show normal circadian photoentrainment. **a–c**, Representative actograms from control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals under a series of lighting paradigms: **a**, 12:12 h LD cycle, 'jet lag', constant darkness (DD), and constant light (LL); **b**, ultradian 3.5:3.5 h cycles; **c**, skeleton photoperiod. The grey background indicates darkness and the yellow dot indicates the 15-min light pulse at CT16 (circadian time 16). *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals have similar photoentrainment to controls with minor deficits in period lengthening. **d**, Percent activity in the dark portion of the ultradian cycle shows no significant difference between the genotypes. **e**, Quantification of phase shifts shows no significant differences between the two groups. **f**, Quantification of circadian period from the two groups under constant dark and constant light conditions. Both groups of animals show significant period lengthening under constant light. **g**, Venn diagram showing Brn3b-positive ipRGCs in yellow and Brn3b-negative ipRGCs in green (full description is provided in Supplementary Table 1). \*\* indicates  $P < 0.01$ , \* indicates  $P < 0.05$  using Student's *t*-test. Error bars represent s.e.m.

However, *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show a minor deficit in period lengthening under constant light conditions (Fig. 4f). Because period lengthening is positively correlated with light intensity, attenuated projections to the IGL in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice could underlie this difference.

Here we show that, although M1 ipRGCs have homogeneous morphological and electrophysiological characteristics, they consist of at least two different subpopulations, which can be discriminated by expression of the Brn3b transcription factor (Fig. 4g). The two M1 subpopulations have distinct brain targets and are involved in different non-image forming visual functions. Using precise molecular genetic tools to ablate Brn3b-expressing ipRGCs, we disrupted the pupillary light reflex but not circadian photoentrainment. Thus, ipRGCs have parallel pathways for controlling non-image forming functions, analogous to the specialized properties of RGCs that mediate image-forming functions<sup>27</sup>.

## METHODS SUMMARY

**Animals.** All experiments were conducted in accordance with NIH guidelines and approved institutional animal care and use committees of the Johns Hopkins University.

**Behavioural analyses.** We used previously described behavioural tests<sup>12</sup> that measure visual acuity (optomotor), pupil constriction (PLR), the period of the circadian oscillator (wheel running activity), the adjustment of the circadian clock to different light stimulations (circadian photoentrainment, 'jet lag' paradigms, phase shifting, and skeleton photoperiod) and direct light effects on activity (constant light and ultradian).

**Histology.** X-gal and alkaline phosphatase histochemistry were performed as described previously<sup>2,5,28</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 March; accepted 18 May 2011.

Published online 17 July 2011.

- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Ecker, J. L. et al. Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010).
- Gooley, J. J., Lu, J., Fischer, D. & Saper, C. B. A broad role for melanopsin in nonvisual photoreception. *J. Neurosci.* **23**, 7093–7106 (2003).
- Hannibal, J. & Fahrenkrug, J. Melanopsin-containing retinal ganglion cells are light responsive from birth. *Neuroreport* **15**, 2317–2320 (2004).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Hattar, S. et al. Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
- Lucas, R. J. et al. Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Panda, S. et al. Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
- Provencio, I., Rollag, M. D. & Castrucci, A. M. Photoreceptive net in the mammalian retina. *Nature* **415**, 493 (2002).
- Ruby, N. F. et al. Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
- Tu, D. C. et al. Physiologic diversity and development of intrinsically photosensitive retinal ganglion cells. *Neuron* **48**, 987–999 (2005).
- Güler, A. D. et al. Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision. *Nature* **453**, 102–105 (2008).
- Hatori, M. et al. Inducible ablation of melanopsin-expressing retinal ganglion cells reveals their central role in non-image forming visual responses. *PLoS ONE* **3**, e2451 (2008).
- Schmidt, T. M., Taniguchi, K. & Kofuji, P. Intrinsic and extrinsic light responses in melanopsin-expressing ganglion cells during mouse development. *J. Neurophysiol.* **100**, 371–384 (2008).
- Wong, K. Y., Dunn, F. A., Graham, D. M. & Berson, D. M. Synaptic influences on rat ganglion-cell photoreceptors. *J. Physiol. (Lond.)* **582**, 279–296 (2007).
- Mrosovsky, N. & Hattar, S. Impaired masking responses to light in melanopsin-knockout mice. *Chronobiol. Int.* **20**, 989–999 (2003).
- Brown, T. M. et al. Melanopsin contributions to irradiance coding in the thalamo-cortical visual system. *PLoS Biol.* **8**, e1000558 (2010).
- Berson, D. M., Castrucci, A. M. & Provencio, I. Morphology and mosaics of melanopsin-expressing retinal ganglion cell types in mice. *J. Comp. Neurol.* **518**, 2405–2422 (2010).
- Hattar, S. et al. Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J. Comp. Neurol.* **497**, 326–349 (2006).

20. Baver, S. B., Pickard, G. E. & Sollars, P. J. Two types of melanopsin retinal ganglion cell differentially innervate the hypothalamic suprachiasmatic nucleus and the olivary pretectal nucleus. *Eur. J. Neurosci.* **27**, 1763–1770 (2008).
21. Badea, T. C., Cahill, H., Ecker, J., Hattar, S. & Nathans, J. Distinct roles of transcription factors Brn3a and Brn3b in controlling the development, morphology, and function of retinal ganglion cells. *Neuron* **61**, 852–864 (2009).
22. Badea, T. C. *et al.* New mouse lines for the analysis of neuronal morphology using CreER(T)/loxP-directed sparse labeling. *PLoS ONE* **4**, e7859 (2009).
23. Schmidt, T. M. & Kofuji, P. Functional and morphological differences among intrinsically photosensitive retinal ganglion cells. *J. Neurosci.* **29**, 476–482 (2009).
24. Mu, X. *et al.* Ganglion cells are required for normal progenitor-cell proliferation but not cell-fate determination or patterning in the developing mouse retina. *Curr. Biol.* **15**, 525–530 (2005).
25. Gan, L., Wang, S. W., Huang, Z. & Klein, W. H. POU domain factor Brn-3b is essential for retinal ganglion cell differentiation and survival but not for initial cell fate specification. *Dev. Biol.* **210**, 469–480 (1999).
26. Lobe, C. G. *et al.* Z/AP, a double reporter for Cre-mediated recombination. *Dev. Biol.* **208**, 281–292 (1999).
27. Wässle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* **5**, 747–757 (2004).
28. Badea, T. C., Wang, Y. & Nathans, J. A noninvasive genetic/pharmacologic strategy for visualizing cell morphology and clonal relationships in the mouse. *J. Neurosci.* **23**, 2314–2322 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Nathans for providing several animal lines (*Brn3b<sup>CKOAP</sup>*, *R26<sup>IAP</sup>* and *Z/AP*) that were crucial for the completion of this study. We thank J. L. Ecker, who created the inducible cre line (*Opn4<sup>CreERT2</sup>*) we used in this study. We thank Z. Yang in D. Zack's laboratory for providing the *Brn3b<sup>Z-dta</sup>* mouse line, which was generously provided by the original laboratory that created this line: W. Klein. We also thank R. Kuruvilla, H. Zhao, M. Halpern, A. P. Sampath and T. Schmidt for their careful reading of the manuscript and helpful suggestions and the Johns Hopkins University Mouse Tri-Lab for support. This work was supported by the National Institutes of Health grant GM076430 (S.H.), the David and Lucile Packard Foundation (S.H.), and the Alfred P. Sloan Foundation (S.H.).

**Author Contributions** S.-K.C., T.C.B. and S.H. performed all experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.H. ([shattar@jhu.edu](mailto:shattar@jhu.edu)) or T.C.B. ([tudor.badea@nih.gov](mailto:tudor.badea@nih.gov)).

## METHODS

**Mice.** All mice were of a mixed background (BL/6;129SvJ). Littermate male animals that were used in the behavioural analyses were aged between 4 and 12 months. Animals were housed and treated in accordance with NIH and IACUC guidelines, and used protocols approved by the Johns Hopkins University Animal Care and Use Committees.

**Generation of *Opn4<sup>CreERT2</sup>* line.** To generate *Opn4<sup>CreERT2</sup>* mice, we used the targeting arms and general strategy detailed in ref. 2. The only difference is that the construct contained a rabbit  $\beta$ -globin intron, CreERT2 recombinase and an IRES LacZ cassette immediately downstream of the start codon for mouse melanopsin.

**Immunohistochemistry.** Mouse retina was fixed as whole eyecup for at least 30 min in 4% paraformaldehyde (PFA) and cryoprotected in 30% sucrose overnight. Retina sections (40  $\mu$ m) were obtained by cryostat and incubated with blocking solution (0.3% Triton X-100 and 5% normal goat serum in PBS) for 1 h before staining with primary antibody overnight at 4 °C. Sections were washed in 1  $\times$  PBS three times for 30 min and incubated with secondary antibody at room temperature for 2 h before mounting in vector-shield mounting solution. Images were taken with an Olympus microscope by epifluorescence. The dilution for melanopsin antibody (Advanced Targeting Systems) is 1:1,000.

**Tamoxifen injections.** The intensity of labelling depends on the amount of tamoxifen injected into animals as well as the efficiency of excision from loxP regions in the reporter mice. Therefore, all intraperitoneal injections of tamoxifen were standardized to label all the identified ipRGC subtypes (M1–M5). In fact, ipRGCs with morphologies characteristic for all identified ipRGC subtypes are observed in the flat mount retinas in Fig. 1b, c. Retina shown in Fig. 1b was from an animal injected with 500  $\mu$ g tamoxifen at P14 (postnatal day 14). Brains shown in Fig. 1d–f were from an animal injected with 250  $\mu$ g tamoxifen at P5. For Fig. 1c, g–i, images are from an animal injected with 1 mg tamoxifen at embryonic day 17 (E17). There is no particular reason to injecting tamoxifen at different developmental times. We simply used the alkaline phosphatase staining as a tracing method to reveal ipRGC targets in the brain.

**Histology.** X-gal staining. Mice were perfused with 15 ml of 4% PFA, the brain was dissected out, cryoprotected in 30% sucrose for 2 days and 50  $\mu$ m coronal sections were obtained by cryostat. Brain sections were incubated in staining solution with 1 mg ml<sup>-1</sup> X-gal for 2 days at room temperature, post-fixed in 4% PFA for 1 h and mounted with glycerol<sup>5</sup>.

Alkaline phosphatase (AP) staining. Mice were perfused with 45 ml of 4% PFA, the brain and retina were dissected out. Whole-mount retina was post-fixed for

30 min and 200  $\mu$ m coronal brain sections were obtained by vibrotome. Both retina and brain sections were heat-inactivated at 65 °C for 90 min and incubated in alkaline phosphatase staining solution. After staining, the sections and retina whole-mount were post-fixed in 4% PFA for overnight and washed with ethanol series before mounting<sup>2, 28</sup>.

**Cholera toxin injections in the eye.** Mice were anaesthetized with avertin. Eyes were injected intravitreally with 2  $\mu$ l of cholera toxin B subunit conjugated with Alexa Fluor 488 (CTB-488) or Alexa Fluor 594 (CTB 594) (Invitrogen). Three days after injection, brains were isolated, sectioned and mounted.

**Visual acuity.** A virtual cylinder OptoMotry (Cerebral Mechanics) was used to determine visual acuity by measuring the image-tracking reflex of mice. A sine-wave grating was projected on the screen rotating in a virtual cylinder. The animal was assessed for a tracking response on stimulation for about 5 s. All acuity thresholds were determined by using the staircase method with 100% contrast.

**Pupillary light reflex.** All animals were kept under 12:12 LD cycle before testing PLR. Before each experiment, all animals were dark-adapted for at least 1 h. While one eye received light stimulation with specific intensity described in the main text from a 470-nm light-emitting-diode light source (Super Bright LEDs), a digital camcorder (DCRHC96; Sony) was used to record from the other eye (for 30 s) at 30 frames per second under a 940-nm light (LDP). The percentage pupil constriction was calculated as the percentage of pupil area at 30 s after initiation of the stimulus (steady state) relative to the dilated pupil size (right before light stimulation). The same group of animals were used for wheel-running activity. The control animals are littermates to the experimental animals (*Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>*) with either *Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>* or *Opn4<sup>+/+</sup>; Brn3b<sup>Z-dta/+</sup>* genotypes.

**Wheel-running activity.** Mice were placed in cages with a 4.5-inch running wheel, and their activity was monitored with VitalView software (MiniMitter). The period was calculated with ClockLab (Actimetrics). Mice were initially placed under 12:12 LD cycle for 2 weeks. Animals were then exposed to two “jet lag” light paradigms: 10 days of a 6-h advance followed by 10 days of a 6-h delay. After the “jet lag” paradigms, mice were kept under constant darkness for 2 weeks followed by 10 days of constant light. Phase-shifting experiments were carried out on the 7th day of constant darkness where each animal was exposed to a 15 min light pulse at CT16 (1,500 lx). Animals were re-entrained to 12:12 LD cycle for 2 weeks before exposing them to ultradian 3.5:3.5 light/dark cycles. The intensity of light for all the light dark cycle were ~1,000 lx. Another set of mice was tested using a skeleton photoperiod, where two 1-h light pulses (800 lx) separated by 10 h of dark were administered.



# HIV-1 adaptation to NK-cell-mediated immune pressure

Galit Alter<sup>1\*</sup>, David Heckerman<sup>2\*</sup>, Arne Schneidewind<sup>1\*</sup>, Lena Fadda<sup>1\*</sup>, Carl M. Kadie<sup>2</sup>, Jonathan M. Carlson<sup>2</sup>, Cesar Oniangue-Ndza<sup>1</sup>, Maureen Martin<sup>3</sup>, Bin Li<sup>1</sup>, Salim I. Khakoo<sup>4</sup>, Mary Carrington<sup>1,3</sup>, Todd M. Allen<sup>1</sup> & Marcus Altfeld<sup>1</sup>

Natural killer (NK) cells have an important role in the control of viral infections, recognizing virally infected cells through a variety of activating and inhibitory receptors<sup>1–3</sup>. Epidemiological and functional studies have recently suggested that NK cells can also contribute to the control of HIV-1 infection through recognition of virally infected cells by both activating and inhibitory killer immunoglobulin-like receptors (KIRs)<sup>4–7</sup>. However, it remains unknown whether NK cells can directly mediate antiviral immune pressure *in vivo* in humans. Here we describe KIR-associated amino-acid polymorphisms in the HIV-1 sequence of chronically infected individuals, on a population level. We show that these KIR-associated HIV-1 sequence polymorphisms can enhance the binding of inhibitory KIRs to HIV-1-infected CD4<sup>+</sup> T cells, and reduce the antiviral activity of KIR-positive NK cells. These data demonstrate that KIR-positive NK cells can place immunological pressure on HIV-1, and that the virus can evade such NK-cell-mediated immune pressure by selecting for sequence polymorphisms, as was previously described for virus-specific T cells and neutralizing antibodies<sup>8</sup>. NK cells might therefore have a previously underappreciated role in contributing to viral evolution.

We hypothesized that HIV-1 can evade recognition by NK cells through the selection of sequence polymorphisms in regions targeted by KIRs, and that KIR-associated polymorphisms in the HIV-1 sequence can be identified on a population level. To test this hypothesis, we examined the relationship between KIR genotypes and HIV-1 polymorphisms in a cohort of 91 untreated, chronically HIV-1-infected individuals (Supplementary Table 1), in whom full-length HIV-1 sequences were determined and HLA-class-I-associated polymorphisms were previously described<sup>9</sup>. We used a decision-tree approach that corrects for phylogenetic structure among the sequences and allows for a multivariate analysis to identify KIR-associated sequence polymorphisms<sup>10</sup>. This analysis led to the identification of 22 positions in the HIV-1 genome at which amino-acid polymorphisms were significantly associated with the presence of a specific KIR gene (Table 1 and Supplementary Fig. 1). Taken together, these data show that HIV-1 can adapt to host KIR genotypes on a population level.

To assess the consequences of these KIR-associated amino-acid polymorphisms for NK-cell-mediated recognition of cells infected with HIV-1, we initially evaluated polymorphisms in a region of HIV-1 that encodes an overlapping segment that spans the carboxy-terminal end of Vpu and the amino-terminal end of Env (polymorphisms 15–18 in Table 1 and Supplementary Fig. 1). We selected this region because these polymorphisms were present in both reading frames at significantly higher frequencies in individuals that possessed at least one copy of the *KIR2DL2* gene than in individuals that did not (Table 2 and Supplementary Tables 2 and 3). In amino-acid positions 71 and 74 of Vpu, HIV-1 sequences derived from *KIR2DL2*-positive individuals encoded a methionine in position Vpu(71) and a histidine in position

Vpu(74) in more than 70% of cases. Because of the overlapping Vpu/Env coding region, these positions of Vpu corresponded to a tryptophan in position Env(17) and a methionine in position Env(20), respectively. The Vpu(71M/74H) (Env(17W/20M)) sequence was significantly less frequent in HIV-1-infected individuals that did not encode *KIR2DL2* ( $P < 0.0001$ ). Furthermore, the presence of the Vpu(71M) polymorphism was in strong linkage disequilibrium with the Vpu(74H) polymorphism ( $P = 3.17 \times 10^{-12}$ ). Taken together, these data demonstrate a significant enrichment of HIV-1 viruses containing the Vpu(71M/74H) (Env(17W/20M)) polymorphism in individuals encoding *KIR2DL2*.

To determine the functional consequences of these *KIR2DL2*-associated polymorphisms in Vpu and Env, we constructed HIV-1 viral variants using the backbone of the HIV-1 NL4-3 strain<sup>11,12</sup>. These variants encoded either the Vpu(71M/74H) (Env(17W/20M)) sequence that was seen in *KIR2DL2*<sup>+</sup> individuals (referred to as Vpu-Env<sup>V/V</sup>, with 'V' standing for variant), or the Vpu(71R/74L) (Env(17G/20L)) sequence that was most commonly seen in *KIR2DL2*<sup>−</sup> subjects (referred to as Vpu-Env<sup>WT/WT</sup>). No significant differences were observed in the ability of the viral variants to replicate in primary CD4<sup>+</sup> T cells *in vitro* (Supplementary Fig. 2a). We subsequently assessed whether the different viral variants had any impact on NK-cell recognition and/or antiviral activity. Primary CD4<sup>+</sup> T cells were infected with viruses containing either the Vpu-Env<sup>V/V</sup> or the Vpu-Env<sup>WT/WT</sup> sequence, and then placed in co-culture with autologous NK

**Table 1 | KIR footprints in HIV-1 sequence**

	Protein	Amino-acid position	KIR association	Consensus amino acid*	Q value
1	Gag	93	KIR2DS3	E	0.1389
2	Gag	138	KIR2DL2	L	0.1852
3	Gag	138	KIR2DS2	L	0.1852
4	Gag	371	KIR2DS5	T	0
5	Gag	389	KIR3DS1	T	0
6	Gag	479	KIR2DS1	I	0
7	Vpr	37	KIR2DS3	I	0.0909
8	Tat	3	KIR2DL2	S	0.0246
9	Tat	3	KIR2DS2	S	0
10	Tat	3	KIR3DS1	S	0.1311
11	Tat	9	KIR2DS3	P	0.1311
12	Tat	28	KIR2DS1	V	0.1311
13	Tat	28	KIR2DS5	V	0.1339
14	Vpu	3	KIR2DL3	S	0.0833
15	Vpu	71	KIR2DL2	M	0.125
16	Vpu	74	KIR2DL2	H	0.1354
17	Env	17	KIR2DL2	W	0
18	Env	20	KIR2DL2	M	0.1667
19	Env	46	KIR3DS1	K	0.1667
20	Env	347	KIR2DS1	L	0.2
21	Env	595	KIR2DS1	I	0.2
22	Nef	9	KIR2DL2	S	0.0833

\* HIV-1 consensus sequence in the 91 study subjects.

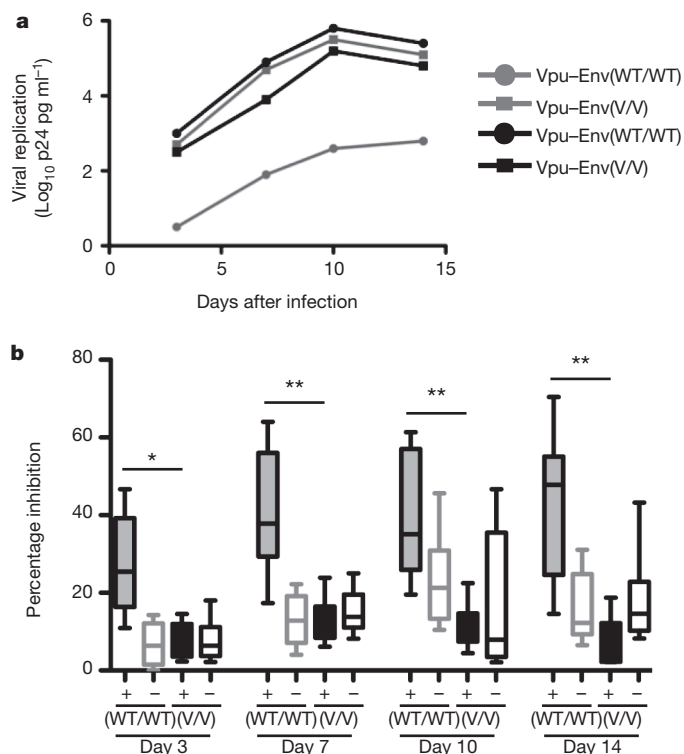
<sup>1</sup>Ragon Institute at MGH, MIT and Harvard, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02129, USA. <sup>2</sup>Microsoft Research, Redmond, Washington 98053, USA. <sup>3</sup>Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC Frederick Inc., NCI-Frederick, Frederick, Maryland 21702, USA. <sup>4</sup>Division of Medicine, Imperial College London, London W2 1PG, UK.

\*These authors contributed equally to this work.

**Table 2 | Frequency of amino-acid polymorphisms among KIR2DL2<sup>+</sup> and KIR2DL2<sup>-</sup> subjects**

KIR2DL2 status	Vpu (amino-acid position)										Percentage of individuals	Number of individuals
	E (70)	<b>M (71)</b>	G (72)	H (73)	<b>H (74)</b>	A (75)	P (76)	W (77)	D (78)	V (79)		
KIR2DL2 <sup>+</sup>	·/P/D	·	·	·	·	·	·	·	N/·	I/·	72	(34/47)
	·	<b>L/Q/·/·</b>	·	·	<b>D/L/·</b>	·	·	·	·	I/·	24	(11/47)
	·	<b>R</b>	·	·	<b>L</b>	·	·	·	·	I/·	4	(2/47)
KIR2DL2 <sup>-</sup>	·/D	·	·/E	·/Q	·	·/D	·	·	·	·	32	(14/44)
	·/H/M	<b>N/V/Q/G</b>	·/A	·	<b>D/R</b>	·	·/L	·/G	·/V	·/I	34	(15/44)
	·/H	<b>R</b>	·	·/R	<b>L</b>	·	·/L	·/G/R/L	·/V	·/I	34	(15/44)
KIR2DL2 status	Env (amino-acid position)										Percentage of individuals	Number of individuals
	R (16)	<b>W (17)</b>	G (18)	T (19)	<b>M (20)</b>	L (21)	L (22)	G (23)	M (24)	L (25)		
KIR2DL2 <sup>+</sup>	·/K	·	·	·/I	·	·	·	·	·/L/I	·	85	(40/47)
	·/K	<b>R/·</b>	·	I	<b>T/·</b>	·	·	·	·	·	11	(5/47)
	·/K	<b>G</b>	·	I/A	<b>L</b>	·/F	·	·	·/I	·	4	(2/47)
KIR2DL2 <sup>-</sup>	·/K	·	·	·/I	·	·	·	·	·/I	·	43	(19/44)
	·	<b>R/M</b>	·	·/I	·/V	·	·	·	·/I	·	23	(10/44)
	·/K	<b>G</b>	·	·/I	<b>L</b>	·	·	·	·/I	·	34	(15/44)
KIR2DL2 status	Gag (amino-acid position)										Percentage of individuals	Number of individuals
	I (134)	V (135)	Q (136)	N (137)	<b>L (138)</b>	Q (139)	G (140)	Q (141)	M (142)	V (143)		
KIR2DL2 <sup>+</sup>	·	·	·	·	·	·	·	·	·	·/I	68	(32/47)
	·/V	·	·	·	<b>M/V/A</b>	·	·	·	·	·/I	32	(15/47)
	·	·	·	·	<b>I</b>	·	·	·	·	·	0	(0/47)
KIR2DL2 <sup>-</sup>	·	·	·/R	·	·	·	·	·	·	·	57	(25/44)
	·	·	·/R	·	<b>M/V/A</b>	·	·	·	·	·	25	(11/44)
	·	·	·	·	<b>I</b>	·	·	·	·	·	18	(8/44)
KIR2DL2 status	Nef (amino-acid position)										Percentage of individuals	Number of individuals
	W (5)	S (6)	K (7)	R (8)	<b>S (9)</b>	V (10)	V (11)	G (12)	W (13)	P (14)		
KIR2DL2 <sup>+</sup>	·/X	·	·	·/X	·	·/X	·/X	·/X	·	·/X	75	(35/47)
	·/X	·	·	·/X	<b>E/R/N/C/L/-</b>	·/X	·/X	·/E	·	·/X	23	(11/47)
	·	·	·	·	<b>K</b>	E	N	·	·	S	2	(1/47)
KIR2DL2 <sup>-</sup>	·/C	·	·/R	·/X	·	·/X	·/X	·/X	·	·/X	52	(23/44)
	·/C	·	·/R	·/X	<b>-/P/R/I/G/W</b>	·/X	·/X	·/X	·	·/X	27	(12/44)
	·	·	·	·/X	<b>K</b>	·/X	·/X	·/D	·	·/S	21	(9/44)

The KIR2DL2-associated sequence polymorphisms in Vpu, Env, Gag and Nef are highlighted in bold. The consensus sequence in the 91 studied individuals is listed in the respective headings. The most common amino acid substitutions are shown for the respective positions. X, positions in which several different amino acids were observed; ·, amino-acid deletions; ·, identical amino acids as in consensus sequence.



cells derived from KIR2DL2<sup>+</sup> or KIR2DL2<sup>-</sup> HIV-1-negative individuals (Fig. 1). Both viruses replicated well in the presence of KIR2DL2<sup>-</sup> NK cells, with less than half a log inhibition of viral replication in the presence of NK cells compared to replication in CD4<sup>+</sup> T cells alone (Fig. 1a, b). In contrast, the Vpu-Env<sup>WT/WT</sup> virus was markedly inhibited by NK cells derived from KIR2DL2<sup>+</sup>, but not KIR2DL2<sup>-</sup>, subjects (Fig. 1b). Thus, the Vpu-Env<sup>V/V</sup> virus, which contains polymorphisms that were strongly associated with the presence of KIR2DL2 on the population level, was not inhibited by KIR2DL2<sup>+</sup> NK cells *in vitro*, whereas the 'wild-type' variant that was rarely observed in KIR2DL2<sup>+</sup> individuals was strongly inhibited by KIR2DL2<sup>+</sup> NK cells, consistent with the selection of Vpu-Env<sup>V/V</sup> viruses in KIR2DL2<sup>+</sup> individuals.

To elucidate further the mechanism by which KIR2DL2<sup>+</sup> NK cells inhibit replication of Vpu-Env<sup>WT/WT</sup> viruses, we monitored the induction of CD107a expression on NK cells *in vitro* after stimulation with autologous CD4<sup>+</sup> T cells infected with either the Vpu-Env<sup>WT/WT</sup>

**Figure 1 | KIR2DL2-associated sequence polymorphisms result in a loss of inhibition of HIV replication by NK cells *in vitro*.** **a**, The Vpu-Env<sup>WT/WT</sup> virus was inhibited more robustly than the Vpu-Env<sup>V/V</sup> virus by NK cells derived from a KIR2DL2<sup>+</sup> individual (grey lines). NK cells derived from a KIR2DL2<sup>-</sup> individual (black lines) did not inhibit either virus. **b**, The Vpu-Env<sup>WT/WT</sup> virus (dark grey bars) was inhibited significantly more strongly than the Vpu-Env<sup>V/V</sup> virus (black bars) by NK cells derived from individuals that expressed KIR2DL2 ( $n = 6$ ). NK cells derived from individuals that did not express KIR2DL2 (white bars,  $n = 6$ ) did not significantly inhibit either virus. \*,  $P < 0.05$ ; \*\*,  $P < 0.005$ ; +, KIR2DL2-positive; -, KIR2DL2-negative. All results are given as mean and s.e.m.

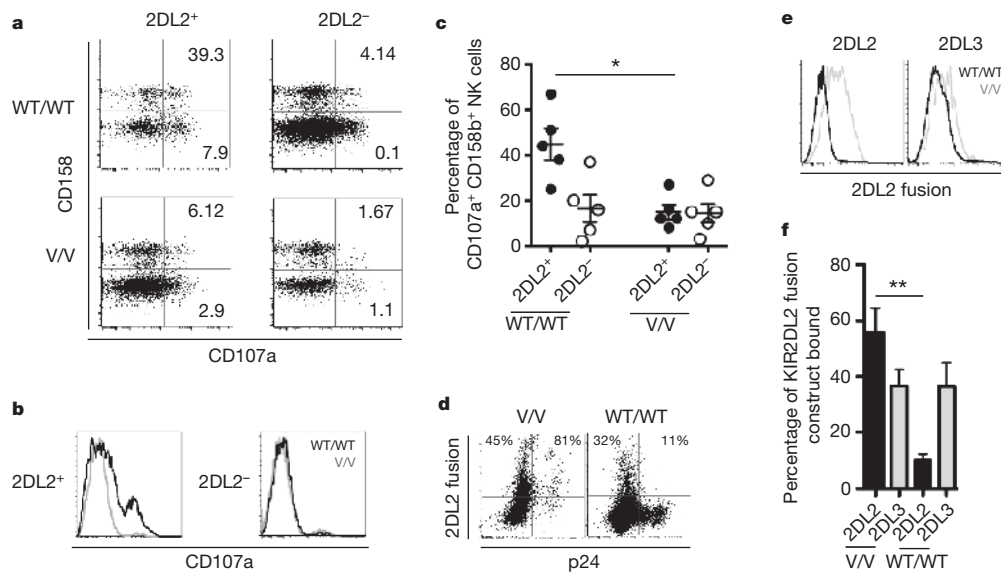
or the Vpu-Env<sup>V/V</sup> virus (Fig. 2a–c). Consistent with the viral inhibition data, CD158b<sup>+</sup> NK cells from KIR2DL2<sup>+</sup> individuals were strongly activated by CD4<sup>+</sup> T cells infected with the Vpu-Env<sup>WT/WT</sup> virus, but not by CD4<sup>+</sup> T cells infected with the Vpu-Env<sup>V/V</sup> virus (Fig. 2a–c). These data are consistent with a model in which the inhibitory NK-cell receptor KIR2DL2 does not bind to cells infected with HIV-1 strains containing the Vpu-Env<sup>WT/WT</sup> sequence, but can bind to cells infected with HIV-1 Vpu-Env<sup>V/V</sup>, providing a strong inhibitory signal to KIR2DL2<sup>+</sup> NK cells and thereby protecting cells infected with Vpu-Env<sup>V/V</sup> viruses from lysis by NK cells.

KIR2DL2 segregates with KIR2DL3, as an allele of the same locus<sup>13</sup>, and is in strong linkage disequilibrium with KIR2DS2 ( $W_n$ , weighted normalized statistic for linkage disequilibrium = 0.976,  $P < 0.001$ ), therefore most individuals in our cohort expressed both these KIRs. However, the three individuals in our cohort who expressed KIR2DL2 in the absence of KIR2DS2 also encoded the Vpu(71M) polymorphism, indicating that KIR2DL2, not KIR2DS2, is responsible for the association with this polymorphism. To test whether KIR2DL2 was directly involved in the recognition of the viral variants, a KIR2DL2–IgG fusion construct was used to assess whether the Vpu-Env<sup>V/V</sup> polymorphism modulated the interaction of KIR2DL2 with HIV-1-infected CD4<sup>+</sup> T cells. The KIR2DL2 fusion construct bound robustly to all uninfected CD4<sup>+</sup> T cells, and also bound to CD4<sup>+</sup> T cells infected with the HIV-1 Vpu-Env<sup>V/V</sup> variant significantly better than to CD4<sup>+</sup> T cells from the same donor infected with the Vpu-Env<sup>WT/WT</sup> variant (Fig. 2d–f). In contrast, the binding of a KIR2DL3 fusion construct to HIV-1-infected CD4<sup>+</sup> T cells was not significantly affected by the KIR2DL2-associated polymorphism (Fig. 2e, f). HLA-C group 1 and group 2 molecules serve as the ligands for the inhibitory receptor KIR2DL2, which has been shown to bind with greater affinity to HLA-C group 1 than group 2 (ref. 13). Staining with the KIR2DL2 fusion construct was consistent with these results, because it bound most strongly to cells from individuals homozygous for HLA-C group 1 (Supplementary Fig. 3). Consistent with these binding data, the Vpu-Env<sup>V/V</sup> polymorphisms were

significantly enriched in KIR2DL2<sup>+</sup> individuals homozygous for HLA-C group 1 ( $P = 0.008$  for Vpu(71M) and  $P = 0.01$  for Vpu(74H)). Taken together, these data indicate that the Vpu-Env<sup>V/V</sup> polymorphism enhances the ability of the inhibitory receptor KIR2DL2 to bind to HIV-1-infected cells, in particular those expressing the ligands with highest affinity for KIR2DL2.

In addition to the Vpu(71M/74H) polymorphism, two additional amino-acid polymorphisms (Gag(138L/I) and Nef(9S/K)) were associated with the presence of the KIR2DL2 gene in the study population (Table 1, polymorphisms 2 and 22). We next determined whether these KIR2DL2-associated polymorphisms were also associated with differential recognition of HIV-1-infected cells by KIR2DL2<sup>+</sup> NK cells. Viruses containing the polymorphisms replicated similarly in CD4<sup>+</sup> T cells (Supplementary Fig. 2b). As observed for the Vpu and Env variants, viruses containing the amino acids that were rarely observed in KIR2DL2<sup>+</sup> individuals (Gag<sup>WT</sup> and Nef<sup>WT</sup>, Table 2 and Supplementary Tables 4 and 5) were inhibited more strongly by NK cells derived from KIR2DL2<sup>+</sup> individuals than were viruses containing the variants selected in KIR2DL2<sup>+</sup> subjects (Fig. 3a, c). Furthermore, KIR2DL2<sup>+</sup> NK cells degranulated more robustly in response to CD4<sup>+</sup> T cells infected with the Gag<sup>WT</sup> and Nef<sup>WT</sup> viruses (Fig. 3b, d), and KIR2DL2–IgG fusion constructs bound significantly less to CD4<sup>+</sup> T cells infected with the Gag<sup>WT</sup> and Nef<sup>WT</sup> viruses than to cells infected with Gag<sup>V</sup> and Nef<sup>V</sup> viruses (Fig. 4a–c). Taken together, these data demonstrate that HIV-1 may evolve in KIR2DL2<sup>+</sup> individuals to enrich particular amino-acid polymorphisms, in an effort to escape recognition by KIR2DL2<sup>+</sup> NK cells.

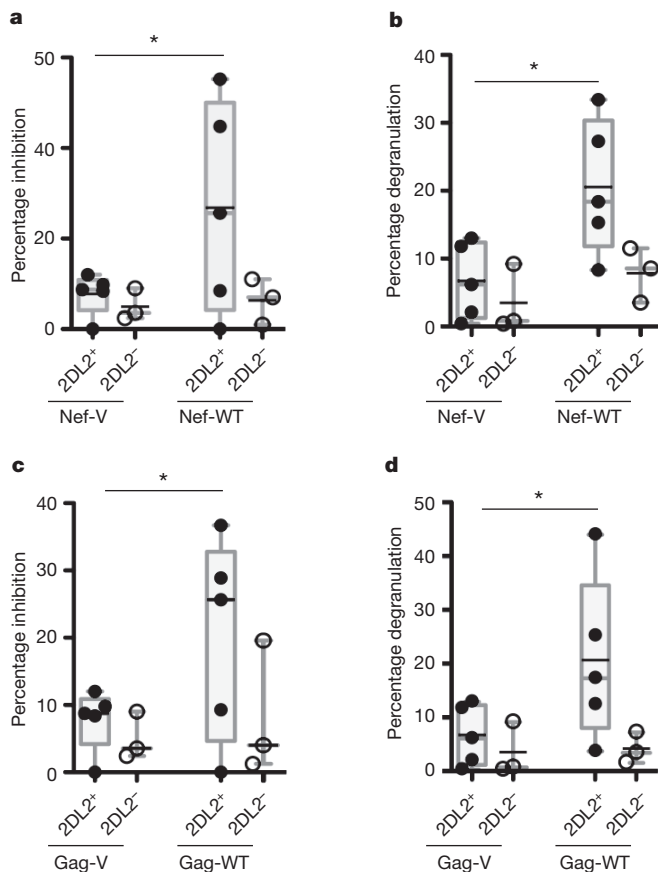
Increasing amounts of evidence indicate that NK cells have an important role in the control of HIV-1 infection<sup>4–6</sup>. Here we report several amino-acid polymorphisms within the HIV-1 clade B sequence that are significantly associated with the expression of specific KIR genes on the population level. We demonstrate in functional studies that these ‘KIR footprints’ can modulate the interaction of KIR<sup>+</sup> NK cells with HIV-1-infected CD4<sup>+</sup> T cells. The selection of particular



**Figure 2 | Amino-acid polymorphisms at positions 71 and 74 in Vpu inhibit KIR2DL2, but not KIR2DL3, recognition and binding.** **a**, Flow cytometric plots depicting the percentages of total CD158b(KIR2DL2/2DL3/2DS2)<sup>+</sup> NK cells that degranulated after co-culture with autologous CD4<sup>+</sup> T cells infected with the Vpu-Env<sup>WT/WT</sup> virus or the Vpu-Env<sup>V/V</sup> virus for two representative subjects (left panels, KIR2DL2<sup>+</sup> subjects; right panels, KIR2DL2<sup>−</sup> subjects). **b**, The percentage of degranulating CD158b<sup>+</sup> NK cells in the total CD158b<sup>+</sup> NK-cell population is also represented in histograms for both the KIR2DL2<sup>+</sup> donor (left panel) and KIR2DL2<sup>−</sup> donor (right panel), for both the WT/WT virus (black line) and the V/V virus (grey line). **c**, Combined data for NK-cell

degranulation in KIR2DL2<sup>+</sup> ( $n = 5$ ) and KIR2DL2<sup>−</sup> ( $n = 5$ ) individuals. \*,  $P < 0.05$ . **d**, Staining pattern of a KIR2DL2–IgG fusion construct on HIV-1-infected CD4<sup>+</sup> T cells from an HLA-C1/C2 heterozygous donor. Percentages indicate the percentage of p24<sup>−</sup> or p24<sup>+</sup> cells that were stained with the KIR2DL2–IgG fusion construct. **e**, Staining of KIR2DL2–IgG and KIR2DL3–IgG fusion constructs on the same donors infected with the WT/WT and V/V viral variants. **f**, Summary of binding data for the KIR2DL2–IgG (black) and KIR2DL3–IgG (grey) fusion constructs for five different HLA-C1 heterozygous CD4<sup>+</sup> T-cell donors after infection with the two viral variants. \*\*,  $P < 0.005$ . All results are given as mean and s.e.m.

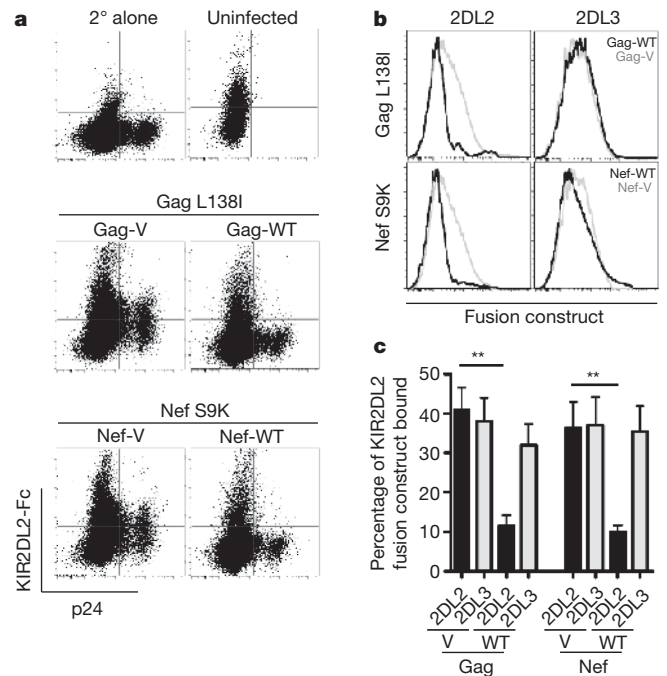




**Figure 3 | Two additional KIR2DL2-associated amino-acid polymorphisms reduce KIR2DL2-mediated NK-cell recognition of virally infected cells.**

**a, c,** NK cells from KIR2DL2<sup>+</sup> individuals ( $n = 5$ ) inhibited the replication of the Nef-WT virus and the Gag-WT virus significantly better than they inhibited replication of the Nef-V and Gag-V viruses. **b, d,** NK cells derived from KIR2DL2<sup>+</sup> individuals ( $n = 5$ ) were activated significantly more by cells infected with the Nef-WT and Gag-WT viruses than by cells infected with Nef-V and Gag-V viruses. However, NK cells derived from KIR2DL2<sup>-</sup> individuals ( $n = 3$ ) did not inhibit viral replication or degranulate in response to cells infected with either of the wild-type or variant viruses. \*,  $P < 0.05$ . All results are given as mean and s.e.m.

amino-acid residues that result in enhanced binding of inhibitory KIRs to infected cells represents a novel approach by which HIV-1 can evade NK-cell-mediated immunity. The molecular mechanisms and precise receptor–ligand interactions involved in this evasion of NK-cell recognition require further investigation. Previous *in vitro* studies have demonstrated that sequence variations in HLA-class-I-presented epitopes<sup>14–23</sup> and small changes in the peptide repertoire presented on HLA class I molecules<sup>24</sup> can both modulate the binding of KIR, providing a potential mechanism for virus-sequence-dependent recognition of infected cells by NK cells. The KIR2DL2-associated sequence polymorphisms studied here had no impact on KIR2DL2 binding to transporter associated with antigen processing (TAP)-deficient T2 cells expressing HLA-Cw\*0102 (Supplementary Fig. 4), despite some degree of HLA-Cw\*0102 stabilization. However, several different HLA-C group 1 molecules might present epitopes in these regions of HIV-1, resulting in differential recognition by KIR2DL2. Additionally, other mechanisms besides the modulation of KIR binding to HLA class I might account for the observed reduction in recognition of variant-virus-infected cells by KIR2DL2<sup>+</sup> NK cells. KIR-associated sequence polymorphisms in HIV-1 proteins might directly modulate the ability of these proteins to be processed and presented<sup>25</sup>, might subtly alter hydrostatic interactions with KIR, or might change the



**Figure 4 | KIR2DL2-associated amino-acid polymorphisms affect binding of KIR2DL2, but not KIR2DL3, to infected CD4<sup>+</sup> T cells.** CD4<sup>+</sup> T cells infected with variant viruses were stained with KIR2DL2-IgG and KIR2DL3-IgG fusion constructs. **a, b,** The dot plots (**a**) and histograms (**b**) show the staining pattern of KIR2DL2-IgG and KIR2DL3-IgG fusion constructs on HIV-1-infected CD4<sup>+</sup> T cells from a HLA-C1/C2 heterozygous donor. **c,** The bar graph summarizes binding data for KIR2DL2 (black) or KIR2DL3 (grey) IgG fusion constructs for five different HLA-C1/C2 heterozygous CD4<sup>+</sup> T-cell donors infected with viral variants. \*\*,  $P < 0.005$ . All results are given as mean and s.e.m.

profile of NK-cell receptor ligands expressed on infected cells, as has been described for HIV-1 Nef, Vpu and Vpr<sup>26,27</sup>.

Protective effects of specific KIRs or combinations of KIR and HLA class I have been described for many infectious diseases, including HIV, hepatitis C virus and human papilloma virus. Although KIR2DL2 has not been identified as a protective KIR gene in any of these viral infections, homozygosity of its allotypic counterpart, KIR2DL3, has been associated with the resolution of hepatitis C infection when KIR2DL3 is co-expressed with its ligand, HLA-C group 1 (ref. 28). This protective KIR2DL3/HLA-C group 1 combination provides a weaker inhibitory signal, resulting in weaker inhibition of NK cells<sup>28</sup>. Here we describe a mechanism by which HIV-1 selects for sequence polymorphisms in KIR2DL2<sup>+</sup> individuals that lead to an enhanced binding of this inhibitory KIR to infected cells, resulting in the inhibition of NK-cell function and thereby enabling HIV-1 to escape the potential protective role of this KIR. Overall, the data from different viral infections are consistent with a model in which enhanced NK-cell activity can contribute to the control of viral replication, and indicate that viruses can evade this NK-cell-mediated immune pressure by selecting for variants that modulate the recognition of infected cells by KIR.

## METHODS SUMMARY

**Viral sequencing.** Genomic DNA was extracted from samples of peripheral blood mononuclear cells and nested PCR protocols were used to amplify HIV-1 genomes<sup>9</sup>.

**Phylogenetic analysis of KIR-associated sequence polymorphisms.** A decision-tree approach<sup>10</sup>, followed by adjustment for multiple comparisons, was used to identify KIR-associated sequence polymorphisms.

**Construction of viruses containing sequence polymorphisms.** Mutations of interest were inserted into the HIV-1 NL4-3 backbone using the GeneTailor site-directed mutagenesis system<sup>11,12</sup>.

**Viral inhibition assay.** Viral inhibition assays were performed after infecting CD4<sup>+</sup> T cells with viral constructs, as indicated. The level of viral inhibition was

then calculated as the difference between viral production (p24 Gag) in wells containing autologous NK cells and production in wells containing CD4<sup>+</sup> T cells alone<sup>6</sup>. **NK-cell degranulation assay.** Degranulation of KIR2DL2<sup>+</sup> NK cells was examined by flow cytometry after co-culture of NK cells with autologous CD4<sup>+</sup> T cells infected with the viral construct indicated, in the presence of Golgi-stop and anti-CD107a-PECy5 for 6 h (ref. 29). NK cells were stained with anti-CD3, anti-CD56, anti-CD16 and anti-CD158b (KIR2DL2/2DL3/2DS2) antibodies, and the level of degranulation was assessed as the proportion of CD107a<sup>+</sup> NK cells among the CD158b<sup>+</sup> NK cells.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 14 December 2009; accepted 31 May 2011.

1. Lanier, L. L. Up on the tightrope: natural killer cell activation and inhibition. *Nature Immunol.* **9**, 495–502 (2008).
2. Yokoyama, W. M. Specific and non-specific natural killer cell responses to viral infection. *Adv. Exp. Med. Biol.* **560**, 57–61 (2005).
3. Lanier, L. L. NK cell recognition. *Annu Rev Immunol.* **23**, 225–274 (2004).
4. Martin, M. P. *et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genet.* **31**, 429–434 (2002).
5. Martin, M. P. *et al.* Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature Genet.* **39**, 733–740 (2007).
6. Alter, G. *et al.* Differential natural killer cell mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *J. Exp. Med.* **204**, 3027–3036 (2007).
7. Altfeld, M. & Goulder, P. 'Unleashed' natural killers hinder HIV. *Nature Genet.* **39**, 708–710 (2007).
8. Goulder, P. J. & Watkins, D. I. HIV and SIV CTL escape: implications for vaccine design. *Nature Rev. Immunol.* **4**, 630–640 (2004).
9. Wang, Y. E. *et al.* Protective HLA class I alleles that restrict acute-phase CD8<sup>+</sup> T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *J. Virol.* **83**, 1845–1855 (2009).
10. Carlson, J. M. *et al.* Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS Comput. Biol.* **4**, e1000225 (2008).
11. Schneidewind, A. *et al.* Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J. Virol.* **81**, 12382–12393 (2007).
12. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* **59**, 284–291 (1986).
13. Moesta, A. K. *et al.* Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. *J. Immunol.* **180**, 3969–3979 (2008).
14. Stewart, C. A. *et al.* Recognition of peptide-MHC class I complexes by activating killer immunoglobulin-like receptors. *Proc. Natl Acad. Sci. USA* **102**, 13224–13229 (2005).
15. Stewart-Jones, G. B. *et al.* Crystal structures and KIR3DL1 recognition of three immunodominant viral peptides complexed to HLA-B\*2705. *Eur. J. Immunol.* **35**, 341–351 (2005).
16. Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G. & Sun, P. D. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* **405**, 537–543 (2000).
17. Thananchai, H. *et al.* Cutting edge: allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *J. Immunol.* **178**, 33–37 (2007).
18. Rajagopalan, S. & Long, E. O. The direct binding of a p58 killer cell inhibitory receptor to human histocompatibility leukocyte antigen (HLA)-Cw4 exhibits peptide selectivity. *J. Exp. Med.* **185**, 1523–1528 (1997).
19. Malnati, M. S. *et al.* Peptide specificity in the recognition of MHC class I by natural killer cell clones. *Science* **267**, 1016–1018 (1995).
20. Peruzzi, M., Parker, K. C., Long, E. O. & Malnati, M. S. Peptide sequence requirements for the recognition of HLA-B\*2705 by specific natural killer cells. *J. Immunol.* **157**, 3350–3356 (1996).
21. Mandelboim, O., Wilson, S. B., Vales-Gomez, M., Reyburn, H. T. & Strominger, J. L. Self and viral peptides can initiate lysis by autologous natural killer cells. *Proc. Natl Acad. Sci. USA* **94**, 4604–4609 (1997).
22. Mandelboim, O. *et al.* The binding site of NK receptors on HLA-C molecules. *Immunity* **6**, 341–350 (1997).
23. Fadda, L. *et al.* Common HIV-1 peptide variants mediate differential binding of KIR3DL1 to HLA-Bw4 molecules. *J. Virol.* **85**, 5970–5974 (2011).
24. Fadda, L. *et al.* Peptide antagonism as a mechanism for NK cell activation. *Proc. Natl Acad. Sci. USA* **107**, 10160–10165 (2010).
25. Le Gall, S., Stamegna, P. & Walker, B. D. Portable flanking sequences modulate CTL epitope processing. *J. Clin. Invest.* **117**, 3563–3575 (2007).
26. Ward, J. *et al.* HIV-1 Vpr triggers natural killer cell-mediated lysis of infected cells through activation of the ATR-mediated DNA damage response. *PLoS Pathog.* **5**, e1000613 (2009).
27. Ward, J. *et al.* HIV modulates the expression of ligands important in triggering natural killer cell cytotoxic responses on infected primary T-cell blasts. *Blood* **110**, 1207–1214 (2007).
28. Khakoo, S. I. *et al.* HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* **305**, 872–874 (2004).
29. Alter, G., Malenfant, J. M. & Altfeld, M. CD107a as a functional marker for the identification of natural killer cell activity. *J. Immunol. Methods* **294**, 15–22 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** These studies were supported by National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases grants R01 AI067031 (M.A.) and P01 AI074415 (M.A. and T.M.A.), and by the Doris Duke Charitable Foundation (M.A.). This project was funded in whole or in part with federal funds from the National Cancer Institute (NIH) under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. S.I.K. is a recipient of a Wellcome Trust Senior Clinical Fellowship and M.A. is a Distinguished Clinical Scientist of the Doris Duke Charitable Foundation. We thank Microsoft Research, the Bill & Melinda Gates Foundation, the Mark and Lisa Schwartz Foundation and the Phillip T. and Susan M. Ragon Foundation for their support.

**Author Contributions** G.A. conducted the immunology experiments and L.F. performed the KIR-staining experiments on T2 cell lines. A.S. and C.O.-N. constructed the viral variants. D.H., C.M.K. and J.M.C. performed the data analysis identifying KIR-associated polymorphisms. B.L. and T.M.A. performed the viral sequencing, M.C. and M.M. performed the HLA and KIR typing, and L.F. and S.I.K. provided the KIR fusion construct. G.A. and M.A. planned the studies, prepared the manuscript and supervised the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.A. ([maltfeld@partners.org](mailto:maltfeld@partners.org)).

## METHODS

**Study subjects.** Ninety-one untreated subjects chronically infected with HIV-1 subtype B, for which HLA class I and KIR genotypes were available (Supplementary Table 1), were included in this study<sup>9</sup>. In addition, 100 HIV-1 negative controls were genotyped for KIR and HLA genotypes. Of this large cohort of uninfected controls, 15 subjects that were KIR2DL2<sup>+</sup> (KIR2DL2<sup>+</sup>/KIR2DL3<sup>+</sup>) and 15 subjects that were KIR2DL2<sup>-</sup> (KIR2DL3<sup>+</sup>/KIR2DL3<sup>+</sup>) were enrolled to provide samples for the generation of NK cells and autologous CD4<sup>+</sup> target cells (Supplementary Table 6). For these *in vitro* studies, only individuals that did not encode KIR3DS1 and HLA-B Bw4-80I were selected, because we had previously observed very strong inhibition of HIV-1 replication *in vitro* in individuals with this combined KIR/HLA genotype<sup>6</sup>. All study subjects were enrolled in Boston through the Massachusetts General Hospital, the Lemuel-Shattuck Hospital and the Fenway Community Health Center. The study was approved by the Massachusetts General Hospital Review Board and all subjects gave written informed consent.

**Viral sequencing.** Genomic DNA was extracted from peripheral blood mononuclear cell samples and nested PCR protocols were used to amplify HIV-1 genomes as described previously<sup>9</sup>. Five independent PCR products of each sample were pooled and directly population-sequenced at the Massachusetts General Hospital DNA sequencing core facility using clade B consensus sequencing primers, as previously described<sup>9</sup>.

**HLA class I and KIR typing.** High-resolution HLA class I typing and KIR genotyping were performed as described previously<sup>6</sup>.

**Phylogenetic analysis of KIR-associated sequence polymorphisms.** We used the decision-tree approach<sup>10</sup>, which corrects for phylogenetic structure among the sequences and allows for a multivariate analysis, to identify KIR-associated sequence polymorphisms. All results were adjusted for multiple comparisons (both *P* and *Q* values are assigned to each result). For each protein analysed, a maximum likelihood phylogenetic tree was constructed from the corresponding sequences. For every KIR and HLA gene, amino-acid position and amino acid at that position, we created two generative or directed graphical models of the observed presence or absence of the amino acid in each sequence: one representing the null hypothesis that the observations are generated by the phylogenetic tree alone and the other representing the alternative hypothesis that additional escape or reversion takes place owing to KIR/HLA pressure in the subjects for which the sequences are observed. The likelihood of the observations was then maximized over the parameters of both models using an expectation-maximization algorithm, and a *P* value was computed using a likelihood ratio test based on those likelihoods<sup>10</sup>. To increase power, the tests were binarized, such that the presence or absence of a given KIR or HLA gene was correlated with the presence or absence of a given amino acid. In addition, KIR-polymorphism pairs were analysed only when the actual or expected count in every cell of the corresponding two-by-two contingency table was  $\geq 3$ . For every amino acid at each position, the KIR or HLA gene with the strongest association (and its corresponding *P* value) was added to the list of identified associations. The analysis was then repeated after removing individuals having or possibly having this KIR or HLA gene. This procedure was iterated until no KIR or HLA gene yielded an association with a *P* value less than 0.05. A *Q*-value statistic, estimating the proportion of false positives among the associations identified, was computed for each association by repeating this analysis on null data (generated by permuting the KIR/HLA data). Correction for multiple comparisons was undertaken using both *Q* < 0.05 (estimating 5% false positives) and *Q* < 0.2 (estimating 20% false positives).

**Construction of viruses containing sequence polymorphisms.** The HIV-1 strain NL4-3 was modified to express one or two mutations in *vpu/env*, *nef* or *gag* using the GeneTailor site-directed mutagenesis system (Invitrogen) or the QuikChange Lightning site-directed mutagenesis system (Stratagene)<sup>11,12</sup>. In brief, mutagenesis was performed using 5' oligonucleotide primers Vpu\_M71R-f (CTTGTGGAGATGGGGTGGAAAGGGGGCACCAT (nucleotide (nt) 6279)), Vpu\_M71R-r (TTTCCACCCCATCTCCACAAGTCTGATACCTTCT (nt 6234)), Vpu\_H74L-f (ATGGGGTGGAAATGGGGCACCTTGCTCCTTGG), Vpu\_H74L-r (GGTGCCCCATTTCACCCCATCTCCACAAG (nt 6247)), Vpu\_M71R/H74L-f (ATGGGGTGGAAAGGGGGCACCTTGCTCCTTGG (nt 6247 and nt 6288)) and Vpu\_M71R/H74L-r (GGTGCCCCCTTTCACCCCATCTCCACAAG (nt 6247)); Nef/S9K-f (GTGGTCAAAAAGTAAAGTGATTGGATGGCC (nt 8827))

and Nef/S9K-r (GGCCATCCAATCACTTTACTTTTGGACCAC (nt 8798)); Gag/L138I-f (CCTATAGTGCAGAACATCCAGGGGCAAATGG (nt 1216)) and Gag/L138I-r (CCATTTCGCCCTGGATGTTCTGCACATATAGG (nt 1186)). Mutated nucleotides are underlined and primer positions are numbered according to the numbering of NL4-3 (GenBank accession number AF324493). The complete HIV-1 coding region of the variant proviruses was sequenced on an ABI3730 XL DNA analyser. Propagation of provirus and generation of viral stocks was performed as previously described<sup>11,12</sup>. Although the full-length sequence of the NL4-3 viruses differed from the autologous sequence of the respective study subjects, the areas flanking the *vpu*, *gag* and *nef* sequences studied were identical between the NL4-3 virus and the respective areas of interest, and only differed in the amino acids indicated in Supplementary Tables 2–5.

**Viral inhibition assay.** Viral inhibition assays using NK cells were performed as previously described<sup>6</sup>. CD4<sup>+</sup> T cells that were generated after 4 days in culture with a bi-specific antibody to CD3 and CD8 were infected with laboratory strains containing amino-acid polymorphisms at a multiplicity of infection of 0.01 for 4 h at 37 °C. Cells were washed twice. Equal numbers of CD4<sup>+</sup> T cells were plated at NK-cell/CD4<sup>+</sup> T cell ratios of 10:1, or alone, for 14 days in the presence of 50 U ml<sup>-1</sup> interleukin-2. Supernatant was collected every 3–4 days for quantification of p24 Gag production by enzyme-linked immunosorbent assay (p24 ELISA; Perkin Elmer).

**NK-cell degranulation assay.** To examine whether KIR2DL2<sup>+</sup> NK cells specifically degranulated in response to autologous CD4<sup>+</sup> T cells infected with either of the viral variants, we monitored the level of CD107a upregulation on KIR2DL2<sup>+</sup> or KIR2DL2<sup>-</sup> NK cells<sup>29</sup>. We selected HIV-1-negative donors that expressed KIR2DL2 in the absence of KIR2DS2 (referred to as KIR2DL2<sup>+</sup>), or KIR2DL3 in the absence of KIR2DL2 and KIR2DS2 (referred to as KIR2DL2<sup>-</sup>). An anti-CD158b antibody was used to detect NK cells expressing KIR2DL2, KIR2DL3 and KIR2DS2. The level of degranulation was assessed as the proportion of CD107a<sup>+</sup> NK cells among the CD158b<sup>+</sup> NK cells. NK cells thus purified were co-cultured in the presence of autologous CD4<sup>+</sup> T cells infected *in vitro* with the respective HIV-1 strains for 7 days. Monensin was added to co-cultures on day 7 at 0.3 µg ml<sup>-1</sup>, in the presence of 20 µl CD107a-PE-Cy5, for 6 h. Cells were washed and stained with CD3-Pacific Blue, CD56-PE-Cy7, CD16-APC-Cy7 (BD Biosciences) and CD158b-PE (Beckman Coulter) for 30 min, then washed and fixed in 1% paraformaldehyde until flow cytometric analysis was performed (FACSCalibur; BD Biosciences).

**KIR-IgG fusion construct binding assay.** Differences in the ability of KIR2DL2 or KIR2DL3 to interact with CD4<sup>+</sup> T cells infected with either of the viral variants were ascertained using KIR2DL2-IgG and KIR2DL3-IgG fusion constructs (provided by S. Khakoo and O. Mandelboim). Peripheral blood mononuclear cells were obtained from HLA-C1/C1 homozygotes, HLA-C1/C2 heterozygotes or HLA-C2/C2 homozygotes, and were treated with 0.3 µg of a bispecific CD3/CD8 antibody in the presence of 50 U interleukin-2 per ml of complete medium. After 3 days, the cells were infected with one of the three KIR2DL2-associated variants (Vpu-71/74, Gag-138 or Nef-9), or with wild-type virus, for 2 days. The cells were then collected and stained with 2 µl of the KIR2DL2-IgG or KIR2DL3-IgG for 1 h on ice. The cells were then washed and stained with a secondary allophycocyanin (APC)-conjugated goat anti-human IgG antibody for an additional 20 min on ice. In parallel, infected cells were stained with the anti-human antibody alone to define the background level of staining. All cells were then fixed with 100 µl of Fix A solution (Invitrogen) for 10 min, washed and permeabilized using 100 µl of Perm B (Invitrogen). The cells were then stained for intracellular p24 using the KC-57-RD1 antibody for 20 min on ice, and then washed. Cells were fixed in 1% paraformaldehyde until flow cytometric analysis was performed (FACSCalibur; BD Biosciences).

**HLA class I stabilization assays.** HLA-C stabilization was assessed in  $2 \times 10^5$  T2 cells that were incubated with 0.04 mg ml<sup>-1</sup> peptide, as indicated, overnight at 26 °C. The following day, peptide-pulsed T2 cells were stained with W632-PE (eBioscience) antibody (HLA-A/B/C) or DT9 antibody (HLA-C/E) for 30 min at 4 °C. Cells were then washed in PBS before staining with anti-mouse IgG-PE (Sigma) for 30 min at 4 °C. Cells were washed twice with PBS and fixed in Perm A solution (BD Biosciences). KIR2DL2-IgG binding to peptide-pulsed T2 cells was assessed as previously described<sup>24</sup>.



# DMRT1 prevents female reprogramming in the postnatal mammalian testis

Clinton K. Matson<sup>1,2</sup>, Mark W. Murphy<sup>1</sup>, Aaron L. Sarver<sup>3</sup>, Michael D. Griswold<sup>4</sup>, Vivian J. Bardwell<sup>1,2,3</sup> & David Zarkower<sup>1,2,3</sup>

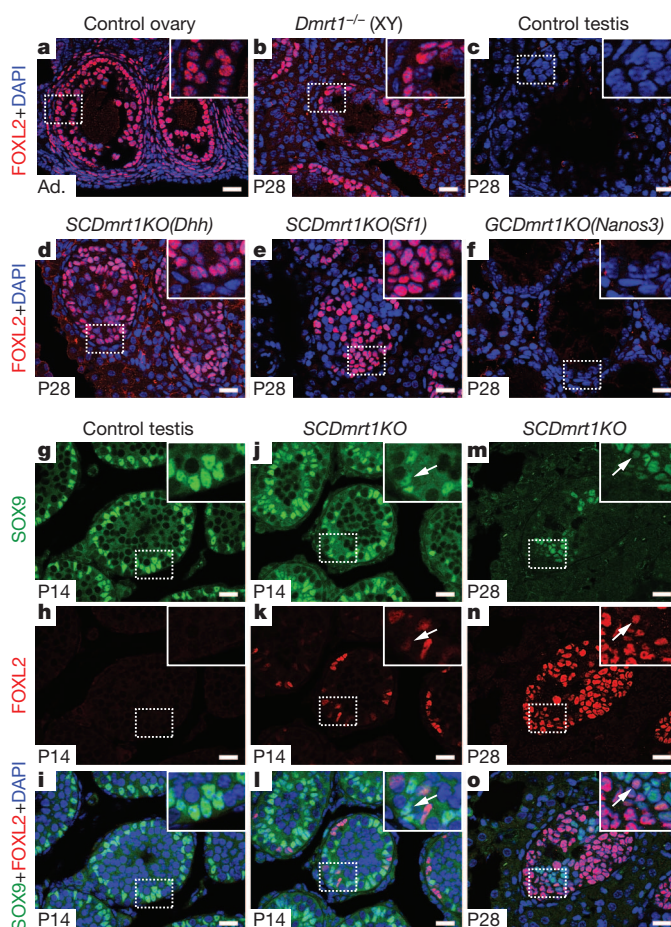
Sex in mammals is determined in the fetal gonad by the presence or absence of the Y chromosome gene *Sry*, which controls whether bipotential precursor cells differentiate into testicular Sertoli cells or ovarian granulosa cells<sup>1</sup>. This pivotal decision in a single gonadal cell type ultimately controls sexual differentiation throughout the body. Sex determination can be viewed as a battle for primacy in the fetal gonad between a male regulatory gene network in which *Sry* activates *Sox9* and a female network involving WNT/ $\beta$ -catenin signalling<sup>2</sup>. In females the primary sex-determining decision is not final: loss of the FOXL2 transcription factor in adult granulosa cells can reprogram granulosa cells into Sertoli cells<sup>2</sup>. Here we show that sexual fate is also surprisingly labile in the testis: loss of the DMRT1 transcription factor<sup>3</sup> in mouse Sertoli cells, even in adults, activates *Foxl2* and reprograms Sertoli cells into granulosa cells. In this environment, theca cells form, oestrogen is produced and germ cells appear feminized. Thus *Dmrt1* is essential to maintain mammalian testis determination, and competing regulatory networks maintain gonadal sex long after the fetal choice between male and female. *Dmrt1* and *Foxl2* are conserved throughout vertebrates<sup>4,5</sup> and *Dmrt1*-related sexual regulators are conserved throughout metazoans<sup>3</sup>. Antagonism between *Dmrt1* and *Foxl2* for control of gonadal sex may therefore extend beyond mammals. Reprogramming due to loss of *Dmrt1* also may help explain the aetiology of human syndromes linked to *DMRT1*, including disorders of sexual differentiation<sup>6</sup> and testicular cancer<sup>7</sup>.

Human chromosome 9p deletions removing *DMRT1* are associated with XY male-to-female sex reversal, and *Dmrt1* homologues determine sex in several non-mammalian vertebrates<sup>8–10</sup>. In mice, *Dmrt1* is expressed and required in both germ cells and Sertoli cells of the testis<sup>11–13</sup>. XY *Dmrt1*-null mutant mice are born as males with testes, although these gonads later undergo abnormal differentiation<sup>14</sup>; hence the role of *Dmrt1* in mammalian sex determination has been unclear (for overview of mammalian sex determination see Supplementary Fig. 1). Here we examine *Dmrt1* mutant testes during postnatal development, asking whether loss of *Dmrt1* causes postnatal feminization in mice.

We first examined gonads of *Dmrt1*-null mutant males (*Dmrt1*<sup>−/−</sup>) for the presence of FOXL2, a female-specific transcription factor expressed in granulosa cells and theca cells<sup>15,16</sup>, the two somatic cell types of the ovarian follicle (Fig. 1a). Four weeks after birth, abundant FOXL2-positive cells were present within mutant seminiferous tubules (Fig. 1b), which in control testes contain only germ cells and Sertoli cells (Fig. 1c). To establish the origin of the FOXL2-positive cells, we deleted *Dmrt1* either in germ cells (using *Nanos3-cre*) or in Sertoli cells (using *Dhh-cre* or *Sf1-cre*) (Supplementary Fig. 2a–l and Supplementary Table 1). Loss of *Dmrt1* in fetal Sertoli cells (*SCDmrt1KO*) but not in fetal germ cells (*GCDmrt1KO*) induced FOXL2 expression (Fig. 1d–f). *SCDmrt1KO* gonads retained small numbers of germ cells, which appeared to arrest in meiotic prophase on the basis of SYCP3 localization (Supplementary Fig. 3). These results demonstrate that DMRT1

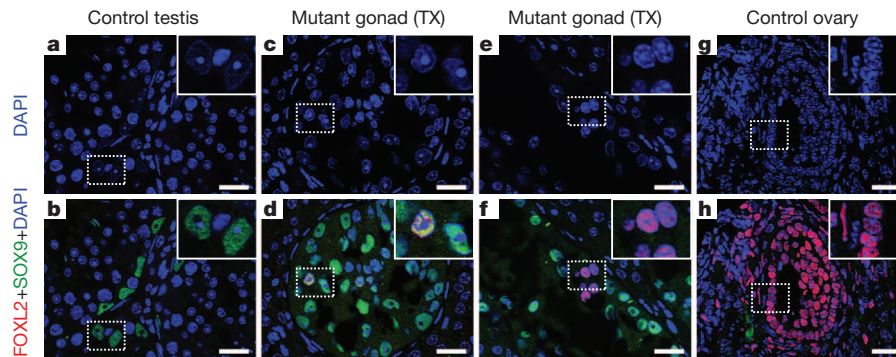
expression in Sertoli cells prevents FOXL2 expression and suggest that *Dmrt1* mutant testes become feminized during the first postnatal month.

Next we examined the timing of FOXL2 induction. At postnatal day (P)7, *SCDmrt1KO* testes had seminiferous tubules in which all Sertoli cells expressed SOX9 normally (Supplementary Fig. 2m–r), but at P14



**Figure 1 | DMRT1 maintains SOX9 and suppresses FOXL2 expression in postnatal Sertoli cells.** a–c, FOXL2 expression detected by immunofluorescence in adult (Ad.) granulosa and theca cells of control ovary (a) and intratubular cells of *Dmrt1*-null testis at P28 (b), but not in control testis (c). DAPI, 4',6-diamidino-2-phenylindole. d–f, FOXL2 is robustly expressed when *Dmrt1* is mutated in fetal Sertoli cells with *Dhh-cre* (d) or *Sf1-cre* (e) but not when *Dmrt1* is mutated in fetal germ cells with *Nanos3-cre* (f). g–o, Timing of FOXL2 expression. FOXL2 is absent from control testis at P14 (g–i). Cells expressing FOXL2 or FOXL2 and SOX9 (arrowheads) are present in *SCDmrt1KO* testis at P14 (j–l). FOXL2-positive cells are abundant in *SCDmrt1KO* testis at P28 and most cells no longer express SOX9 (m–o). Scale bars, 20  $\mu$ m.

<sup>1</sup>Developmental Biology Center and Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis 55455, Minnesota, USA. <sup>2</sup>Molecular, Cellular, Developmental Biology, and Genetics Graduate Program, University of Minnesota, Minneapolis 55455, Minnesota, USA. <sup>3</sup>University of Minnesota Masonic Cancer Center, Minneapolis, Minnesota 55455, USA. <sup>4</sup>School of Molecular Biosciences, Washington State University, Pullman, Washington 99164, USA.



**Figure 2 | Sertoli-to-granulosa transdifferentiation in the adult testis.** **a–h**, Expression of FOXL2 and SOX9 one month after tamoxifen (TX) injection into *Dmrt1*<sup>lox/lox</sup> adult males (8 weeks and older) carrying inducible ubiquitous *cre* transgene *UBC-cre/ERT2*. **a, b**, Sertoli cells in control testis express SOX9 but not FOXL2. **c–f**, Mutant testis has Sertoli-like cells expressing SOX9 or SOX9 and FOXL2 (**d**, inset) and granulosa-like cells expressing only FOXL2

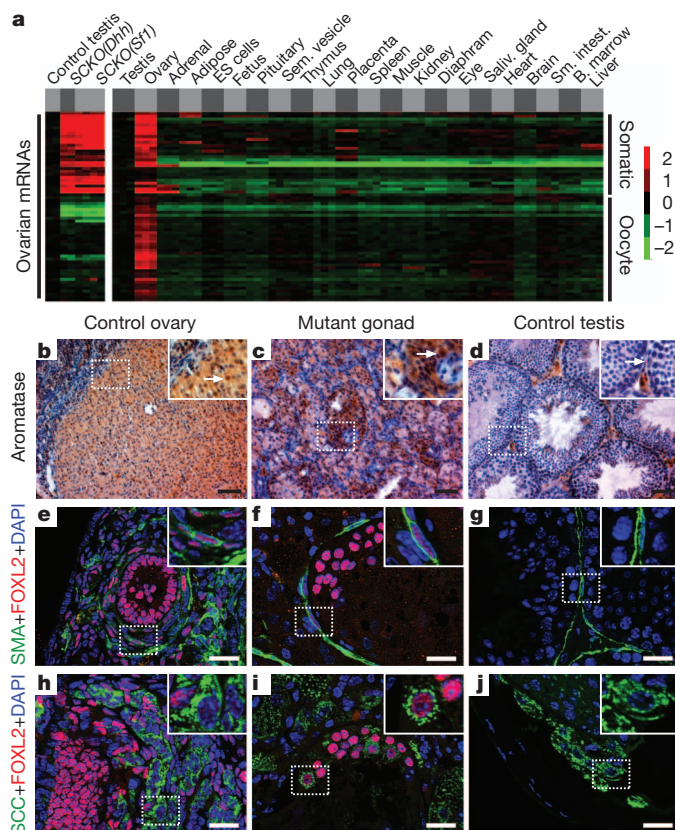
some intratubular cells co-expressed SOX9 and FOXL2 or lacked SOX9 and strongly expressed FOXL2 (Fig. 1g–l). By P28 few SOX9-positive cells remained and most intratubular cells strongly expressed FOXL2 (Fig. 1m–o). Histological analysis of mutant gonads is shown in Supplementary Fig. 4. These results show that fetal loss of *Dmrt1* causes postnatal Sertoli cells to lose the male-promoting SOX9 and instead express the female-promoting FOXL2.

Loss of *Foxl2* in the adult ovary can lead to transdifferentiation of granulosa cells to Sertoli cells<sup>2</sup>, so we asked whether loss of *Dmrt1* in the adult testis activates *Foxl2* and causes the reciprocal sex transformation, from Sertoli to granulosa. Indeed, one month after deletion of *Dmrt1* in adult males (using a tamoxifen-inducible *cre* transgene), we observed cells with typical Sertoli cell features including tripartite nucleoli but expressing both SOX9 and FOXL2 (Fig. 2a–d), as well as

(f, inset). **g, h**, FOXL2-positive cells in control ovary have DAPI morphology similar to FOXL2 single-positive cells of mutant testis. FOXL2-positive cells in mutant testis resemble granulosa cells: they lack the tripartite nucleoli of Sertoli cells, have smaller and more rounded nuclei, and have more punctate DAPI staining. *UBC-cre/ERT2* also deletes *Dmrt1* in germ cells, causing precocious meiosis<sup>12</sup>; after one month germ cells are nearly absent. Scale bars, 20  $\mu$ m.

cells with typical granulosa cell nuclear morphology that lacked SOX9 and strongly expressed FOXL2 (Fig. 2e–h). Thus antagonism between DMRT1 and FOXL2 continues into adulthood and Sertoli cell fate remains plastic even after terminal differentiation.

To evaluate further the transformation of mutant gonads, we compared the messenger RNA profile of control and mutant P28 testes; 5,030 mRNAs were expressed >8-fold differently across this data set or a data set comparing testis to 21 other tissues including ovary (Supplementary Fig. 5a). We calculated Pearson correlation coefficients for expression of these 5,030 mRNAs in mutant gonads relative to each tissue and found that the mutant gonad most closely resembled ovary (Supplementary Fig. 5b; average  $R = 0.75$ ). Many mRNAs with decreased expression in mutant gonads also were low in other tissues, probably reflecting a lack of male germ cells, which comprise much of the testis mass. Also, some mRNAs that were increased in mutant gonads were increased in other tissues. Therefore, to specifically evaluate ovary-enriched mRNAs, we used bioGPS (<http://www.biogps.gnf.org>; see Supplementary Information) to identify 65 mRNAs with expression closely correlated to *Foxl2* and then compared their expression in ovary relative to the other 21 tissues (Fig. 3a and Supplementary Fig. 6). This comparison confirmed that these mRNAs are highly ovary enriched. About 40% were increased in mutant gonads relative to control testes; about 80% of the remainder were oocyte enriched. Thus loss of *Dmrt1* causes large changes in mRNA expression, including induction of multiple ovary-enriched mRNAs. mRNA profiling of *Dmrt1* mutant gonads perinatally and at P9 did not reveal



**Figure 3 | Feminization of *SCDmrt1KO* XY gonads.** **a**, Expression of ovary-enriched mRNAs with expression profiles similar to *Foxl2* (see Supplementary Information). mRNAs labelled 'somatic' were enriched in ovarian somatic cells; those labelled 'oocyte' were enriched in female germ cells. See Supplementary Fig. 6 for higher resolution image. **b**, marrow, bone marrow; Saliv. gland, salivary gland; Sem. vesicle, seminal vesicle; Sm. intestine, small intestine. **b–d**, Immunohistochemistry detection of CYP11A1/aromatase expression in follicles of control adult ovary (**a**) and in adult XY *SCDmrt1KO* gonad (**b**), but only in interstitial Leydig cells of control testis (**c**). Arrows indicate aromatase-positive granulosa cells in ovary and mutant gonad and negative Sertoli cell in control testis. Scale bars, 50  $\mu$ m. **e–g**, Immunofluorescence detection of SMA and FOXL2. Ovarian theca cells (**e**, inset) are elongated cells expressing both proteins; similar cells are present in mutant gonads (**f**); peritubular myoid cells in control testes express SMA and not FOXL2 (**g**). Scale bars, 20  $\mu$ m. **h–j**, Immunofluorescence detection of cells coexpressing FOXL2 in the nucleus and steroidogenic enzyme CYP11A1/SCC at high levels in the cytoplasm in control ovary (**h**) and XY *SCDmrt1KO* gonads (**i**). SCC-positive cells in control testis (**j**) are interstitial Leydig cells. Mutant gonads were *SCDmrt1KO(Dhh)*. All tissues were from adult mice. Scale bars, 20  $\mu$ m.



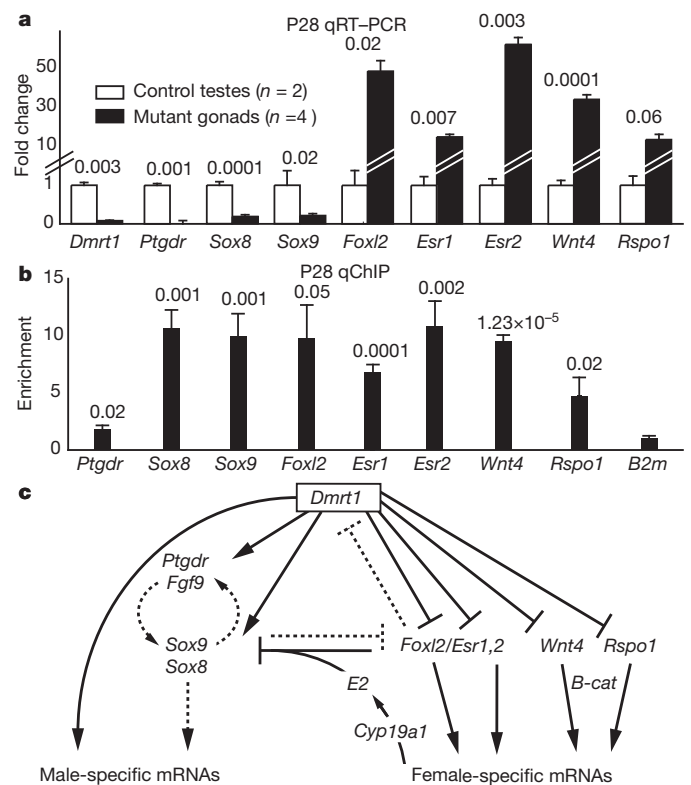
apparent feminization<sup>17,18</sup>, consistent with the observation that FOXL2 expression starts at ~P14.

Further analysis of the mRNA profiling data identified highly increased expression (>5-fold,  $P < 0.001$ ) of many mRNAs expressed in granulosa cells and required for ovarian development or function. These included *Foxl2*, *Nr5a2* (also known as *Lrh1*), *Wnt4*, LH receptor (*Lhcgr*), prolactin receptor (*Prlr*), FSH receptor (*Fshr*), follistatin (*Fst*), *Sfrp4*, *Igfbp5*, *Inhbb*, *Inha* and *Lnfg* (Supplementary Table 2). *Foxl2os*, a noncoding RNA transcribed from the opposite strand of the *Foxl2* coding region, also was highly overexpressed and has been suggested as a positive regulator of *Foxl2* (ref. 19). We confirmed increased expression in mutant gonads of LRH1, a transcription factor expressed only in granulosa cells within the ovary<sup>20</sup> and absent from the testis (Supplementary Fig. 7a–f). *Nr5a2* is probably a direct target of DMRT1 regulation, based on binding of DMRT1 to its promoter proximal sequences *in vivo* (Supplementary Fig. 7g). On the basis of mRNA and protein expression data and changes in cellular morphology, we conclude that loss of *Dmrt1* in testes reprograms Sertoli cells into granulosa cells.

Granulosa cells produce oestrogens, which are essential for ovarian development in many vertebrates; in mammals, oestrogen signalling also acts with FOXL2 to repress *Sox9* transcription in adult granulosa cells<sup>2</sup>. HSD17 $\beta$ 1 and CYP19A1/aromatase are enzymes critical for oestrogen synthesis, and mRNAs for both enzymes were increased in mutant gonads (Supplementary Fig. 8). Aromatase protein is robustly expressed in granulosa cells and was strongly expressed in mutant gonads (Fig. 3b–d). Consistent with these enzyme changes, oestradiol was raised in the serum of adult mutants relative to control adult males (Supplementary Information). Although expression of the androgenic enzyme *Hsd17 $\beta$ 3* was not affected in mutant gonads (Supplementary Fig. 8), androgen levels were reduced based on severely decreased seminal vesicle weight, a sensitive indicator of androgen activity ( $350 \pm 52$  mg versus  $182 \pm 36$  mg;  $n = 3$ ,  $P = 0.01$ ).

Theca cells are induced during follicle growth in the ovary, probably in response to granulosa cell signals<sup>21</sup>, and together with granulosa cells and oocytes they comprise the functional unit of the ovary. Because mutant gonads contained apparently functional granulosa cells, we asked whether theca cells also formed. Theca cells have spindle-shaped nuclei and express both FOXL2 and smooth muscle actin (SMA) (Fig. 3e). Adult mutant gonads contained cells closely resembling theca cells and expressing both proteins (Fig. 3f). The theca-like cells probably derive either from granulosa cells or peritubular myoid cells (which also are elongated and express SMA; Fig. 3g). However, as seminiferous tubule integrity was lost before formation of these cells (Fig. 3f and Supplementary Fig. 9), they could potentially derive from interstitial cells that invaded the tubule remnants. We also observed intratubular cells strongly expressing the steroidogenic enzyme SCC (Fig. 3h–j); these cells resembled luteinized granulosa cells of the ovary (Fig. 3h), suggesting that granulosa cells in the mutant gonad are responsive to gonadotropins. We therefore tested the effect of exogenous gonadotropin stimulation; treated mutants, but not controls, had additional luteinized granulosa cells and germ cells with oocyte-like nuclear morphology that expressed the oocyte-specific proteins MATER and ZP2 (Supplementary Fig. 10). This result indicates that both somatic cells and germ cells are feminized in mutant gonads.

The preceding results indicate that DMRT1 is essential for postnatal sex maintenance. DMRT1 is a sequence-specific transcriptional regulator, capable of activating or repressing transcription of target genes<sup>18,22</sup>. To help find targets of DMRT1 regulation with potential roles in sex maintenance we examined expression of known fetal sex-determining genes in mutant gonads at P28 by quantitative polymerase chain reaction with reverse transcription (qRT–PCR; Fig. 4a). Among masculinizing genes, *Ptgd*, *Sox9* and *Sox8*, which acts redundantly with *Sox9* (refs 23, 24), were reduced. Among feminizing genes, *Foxl2*, *Esr1*, *Esr2*, *Wnt4* and *Rspo1* were increased. We assayed binding of DMRT1 to DNA of P28 testes by quantitative chromatin immunoprecipitation



**Figure 4 | DMRT1 regulation of postnatal gene expression.** **a**, qRT–PCR analysis of sex-determining genes at P28. Significance of expression changes is indicated (Student's *t*-test). Mutant gonads were *SCDmrt1KO* (*Sf1*); *SCDmrt1KO* (*Dhh*) mutant gonads and equivalent expression changes. **b**, qChIP analysis of DMRT1 DNA binding in P28 testes. Significance of enrichment relative to *B2m* (Student's *t*-test) is shown. **c**, Model for regulation by postnatal sex maintenance by DMRT1. Proposed direct regulation based on ChIP and mRNA expression data are indicated by solid lines; indirect or potential regulation is indicated by dashed lines. Model adapted from ref. 2.

(qChIP), guided by genome-wide ChIP data from P9 testes (ChIP-chip<sup>18</sup> and ChIP-seq (unpublished data)). DMRT1 bound both upstream and downstream of *Sox9* and upstream of *Sox8*, and bound weakly near *Ptgd*. DMRT1 bound strongly near *Foxl2*, *Esr1*, *Esr2*, *Wnt4* and *Rspo1* (Fig. 4b). All of the DMRT1-associated regions contained at least one close match to the DMRT1 DNA-binding consensus<sup>18,22</sup>.

On the basis of mRNA and protein expression data and ChIP analysis, we propose a model for postnatal sex maintenance (Fig. 4b) in which DMRT1 maintains male fates by repressing multiple female-promoting genes and activating male-promoting genes. *Sox9* is dispensable for testis differentiation after sex determination<sup>24,25</sup>, suggesting that other critical male regulators remain to be found; *Sox8* is a clear candidate based on its redundancy with *Sox9* (refs 23, 24). We find that DMRT1 represses *Foxl2*, which is known to maintain postnatal ovarian fate. FOXL2 also represses *Dmrt1* (ref. 2); thus antagonism between these sex-specific transcriptional regulators may be central to sex maintenance in both sexes throughout reproductive life. *Wnt4* and *Rspo1* also are prime candidates for postnatal sex maintenance based on their requirement in ovarian determination in the fetus<sup>26,27</sup>. Indeed, P28 mutant gonads had increased nuclear  $\beta$ -catenin in somatic cells, as in ovaries, but control testes did not, indicating active WNT/ $\beta$ -catenin signalling in the mutant gonads (Supplementary Fig. 11). Functional analysis of *Wnt4*, *Rspo1* and other known fetal sex regulators will be important to establish their roles in sex maintenance.

The analysis presented here demonstrates that deletion of *Dmrt1* during fetal development induces postnatal feminization of the testis, causing male-to-female primary sex reversal. Moreover, deletion of



*Dmrt1* in adults can reprogram differentiated Sertoli cells into apparent granulosa cells. Why *Dmrt1* mutants are feminized only after birth remains unclear. Another male-promoting gene may act redundantly with *Dmrt1* before P14, masking its function; alternatively, the testis may lack potential feminizing activity from genes such as *Foxl2* before P14. Another puzzle is that *Dmrt1* mutant mice are born male, whereas human 9p deletions removing *DMRT1* can cause XY feminization at birth. The human sex reversal may reflect failure to maintain male sex determination, and the longer human gestation may permit testis-to-ovary reprogramming before birth. Alternatively, human testes may have potential feminizing activity earlier or may lack masculinizing genes redundant with *DMRT1*. Our results may provide insights into the aetiology of human gonadal disorders, including gonadoblastoma and granulosa cell tumours of the testis. Moreover, because many genes implicated in this study are evolutionarily conserved, similar mechanisms may control adult sex switching in fish and may maintain sexual fate in the adult gonads of other vertebrates or even in other phyla.

## METHODS SUMMARY

**Mouse breeding.** *Dmrt1* mutant and control males were generated as described<sup>12</sup>; tissue-specific Cre recombinase strains are in Supplementary Table 1. Adult wild-type or *Dmrt1*<sup>flx/flx</sup> females were used as controls. Mice were mixed C57BL/6J, 129S1 and FVB genetic background. Protocols were approved by the Institutional Animal Care and Use Committee.

**Immunofluorescence and immunohistochemistry.** Immunofluorescence and immunohistochemistry were performed as described<sup>12</sup>. Antibodies are listed in Supplementary Table 3. Analyses included at least two biological replicates.

**Tamoxifen treatment.** Tamoxifen-inducible deletion of *Dmrt1* in adult males was as described<sup>12</sup>. Testes were harvested one to two months after treatment.

**mRNA expression analysis.** mRNA expression profiling and data analysis were as described<sup>13</sup> except total testis RNA was isolated from 4-week-old mice using TRIzol reagent (Invitrogen no. 15596-026). Additional detail is in Supplementary Methods.

**qRT-PCR.** qRT-PCR was as described<sup>12</sup>. qRT-PCR primers are listed in Supplementary Table 4.

**ChIP.** ChIP followed by either microarray (ChIP-chip) or qPCR analysis (qChIP) were as described<sup>18</sup>. Gene-specific primers used for qChIP are in Supplementary Table 4.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 February; accepted 27 May 2011.

Published online 20 July 2011.

- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121 (1991).
- Uhlenhaut, N. H. *et al.* Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130–1142 (2009).
- Raymond, C. S. *et al.* Evidence for evolutionary conservation of sex-determining genes. *Nature* **391**, 691–695 (1998).
- Loffler, K. A., Zarkower, D. & Koopman, P. Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: *FOXL2* is a conserved, early-acting gene in vertebrate ovarian development. *Endocrinology* **144**, 3237–3243 (2003).
- Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J. & Zarkower, D. Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* **215**, 208–220 (1999).
- Tannour-Louet, M. *et al.* Identification of *de novo* copy number variants associated with human disorders of sexual development. *PLoS ONE* **5**, e15392 (2010).
- Turnbull, C. *et al.* Variants near *DMRT1*, *TERT* and *ATF7IP* are associated with testicular germ cell cancer. *Nature Genet.* **42**, 604–607 (2010).

- Yoshimoto, S. *et al.* A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl Acad. Sci. USA* **105**, 2469–2474 (2008).
- Smith, C. A. *et al.* The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
- Matsuda, M. *et al.* *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**, 559–563 (2002).
- Kim, S., Bardwell, V. J. & Zarkower, D. Cell type-autonomous and non-autonomous requirements for *Dmrt1* in postnatal testis differentiation. *Dev. Biol.* **307**, 314–327 (2007).
- Matson, C. K. *et al.* The mammalian doublesex homolog *DMRT1* is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev. Cell* **19**, 612–624 (2010).
- Krentz, A. D. *et al.* The DM domain protein *DMRT1* is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. *Proc. Natl Acad. Sci. USA* **106**, 22323–22328 (2009).
- Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. & Zarkower, D. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587–2595 (2000).
- Schmidt, D. *et al.* The murine winged-helix transcription factor *Foxl2* is required for granulosa cell differentiation and ovary maintenance. *Development* **131**, 933–942 (2004).
- Uda, M. *et al.* *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* **13**, 1171–1181 (2004).
- Fahrioglu, U., Murphy, M. W., Zarkower, D. & Bardwell, V. J. mRNA expression analysis and the molecular basis of neonatal testis defects in *Dmrt1* mutant mice. *Sex. Dev.* **1**, 42–58 (2007).
- Murphy, M. W. *et al.* Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog *DMRT1* in the juvenile testis. *Proc. Natl Acad. Sci. USA* **107**, 13360–13365 (2010).
- Cocquet, J., Pannetier, M., Fellous, M. & Veitia, R. A. Sense and antisense *Foxl2* transcripts in mouse. *Genomics* **85**, 531–541 (2005).
- Duggavathi, R. *et al.* Liver receptor homolog 1 is essential for ovulation. *Genes Dev.* **22**, 1871–1876 (2008).
- Orisaka, M., Tajima, K., Tsang, B. K. & Kotsuji, F. Oocyte-granulosa-theca cell interactions during preantral follicular development. *J. Ovarian Res.* **2**, 9 (2009).
- Murphy, M. W., Zarkower, D. & Bardwell, V. J. Vertebrate DM domain proteins bind similar DNA sequences and can heterodimerize on DNA. *BMC Mol. Biol.* **8**, 58 (2007).
- Chaboissier, M. C. *et al.* Functional analysis of *Sox8* and *Sox9* during sex determination in the mouse. *Development* **131**, 1891–1901 (2004).
- Barrionuevo, F. *et al.* Testis cord differentiation after the sex determination stage is independent of *Sox9* but fails in the combined absence of *Sox9* and *Sox8*. *Dev. Biol.* **327**, 301–312 (2009).
- Chang, H. *et al.* *Wt1* negatively regulates  $\beta$ -catenin signaling during testis development. *Development* **135**, 1875–1885 (2008).
- Vainio, S., Heikkilä, M., Kispert, A., Chin, N. & McMahon, A. P. Female development in mammals is regulated by *Wnt-4* signalling. *Nature* **397**, 405–409 (1999).
- Parma, P. *et al.* *R-spondin1* is essential in sex determination, skin differentiation and malignancy. *Nature Genet.* **38**, 1304–1309 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Treier for helpful discussion, K. Hatzl, A. Minkina, A. Peterson, the University of Minnesota Mouse Genetics Laboratory and C. Small for technical assistance, J. Dean, R. Veitia and K.-i. Morohashi for antibodies, D. Greenstein and A. M. Weber-Main for comments on the manuscript, C. Manivel for histology expertise, M. Steffes and D. Gabrielson for oestradiol analysis, and the University of Minnesota Supercomputing Institute for computational resources. This work was funded by the NIH (GM59152), the Minnesota Medical Foundation, and a predoctoral fellowship from the NSF (to C.K.M.).

**Author Contributions** C.K.M. performed mouse breeding and analysis of protein and mRNA expression; M.W.M. performed ChIP analysis; A.L.S. performed bioinformatic analysis; C.K.M., D.Z. and V.J.B. designed the study, analysed data, and wrote the paper; M.D.G. provided mRNA profiling expertise; all authors discussed the results and edited the paper.

**Author Information** mRNA expression profiling data have been deposited at the Gene Expression Omnibus under accession number GSE27261 and can be reviewed at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=z1ctbiugsamymtc&acc=GSE27261>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.Z. (zarko001@umn.edu) or V.J.B. (bardw001@umn.edu).

## METHODS

**Mouse breeding.** Conditional *Dmrt1* mutant and control males were generated as described<sup>12</sup>; tissue-specific Cre recombinase strains are in Supplementary Table 1. Adult wild-type or *Dmrt1*<sup>fllox/fllox</sup> females were used as controls. Mice were of mixed C57BL/6J, 129S1 and FVB genetic background. Protocols were approved by the University of Minnesota Institutional Animal Care and Use Committee.

**Immunofluorescence and immunohistochemistry.** Both immunofluorescence and immunohistochemistry were performed as described<sup>12</sup>. Antibodies are listed in Supplementary Table 3. Analyses included a minimum of two biological replicates.

**Tamoxifen treatment.** Tamoxifen-inducible deletion of *Dmrt1* in adult males was performed as previously described<sup>12</sup>. Testes were harvested one to two months after treatment.

**mRNA expression analysis.** mRNA expression profiling and data analysis were performed as described<sup>13</sup> except total testis RNA was isolated from 4-week-old mice using TRIzol reagent (Invitrogen no. 15596-026). Affymetrix Mouse Genome 439 2.0 arrays were normalized by GC-RMA normalization<sup>28</sup> using GeneData Refiner. The Raw .cel files and the normalized data are deposited in the Gene Expression Omnibus (GEO)<sup>29</sup> under accession number GSE27261. GSE9954 was obtained from the GEO database. The arrays with the highest sample identification numbers were removed from the tissue data set to select 22 tissue types, each with three experimental replicates. When multiple probe sets were mapped to the same gene symbol, these values were averaged to obtain one value for each gene symbol. Direct Pearson correlation *R* values were calculated using all array data following reduction to gene symbols, and these values are shown in Fig. 2b.

Each experiment in our data set was divided by the average expression value from control testis tissue. GSE9954 data were separately divided by the average signal obtained from the GSE9954 testis samples. This was done separately for each data set to determine how samples from each data set differed from a baseline 'testis' expression state. Cluster 3.0 software<sup>30</sup> was used to: (1) log base 2 transform

the data; (2) filter the data set for genes that showed at least three observations with  $\text{abs}(\text{val}) > 3$  (eightfold), which resulted in 5,030 genes passing the filter using both data sets combined; and (3) cluster the data on the gene axis using average linkage hierarchical clustering. The experimental axis was defined by order of decreasing correlation to the mutant testes calculated as described earlier. Javatreview Software<sup>31</sup> was used to generate heatmap images.

**qRT-PCR.** qRT-PCR was performed as described<sup>12</sup>. qRT-PCR primers are listed in Supplementary Table 4.

**Chromatin immunoprecipitation.** ChIP followed by either microarray (ChIP-chip) or qPCR analysis (qChIP) were performed as described<sup>18</sup>. Gene-specific primers used for qChIP are in Supplementary Table 4.

**Oestradiol assays.** Serum oestradiol was assayed using a clinical electrochemiluminescence immunoassay (Roche Estradiol II, 03000079 122) according to manufacturer's instructions. Three of three males assayed had levels below the detection limit, whereas two of three females had measurable oestradiol (5.0 and 19.7 pg dl<sup>-1</sup>). Two of three *SCDmrt1KO(Dhh)* mutant males had measurable oestradiol (5.6 and 21.2 pg dl<sup>-1</sup>).

**Gonadotropin treatment.** Six-to-eight-week-old mutant males, control males and control females were treated with 5 units of pregnant mare serum by intraperitoneal injection and gonads were harvested 48 h later.

28. Wu, J., Irazarray, R. A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99**, 909–917 (2004).
29. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
30. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
31. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).

# A Polycomb-based switch underlying quantitative epigenetic memory

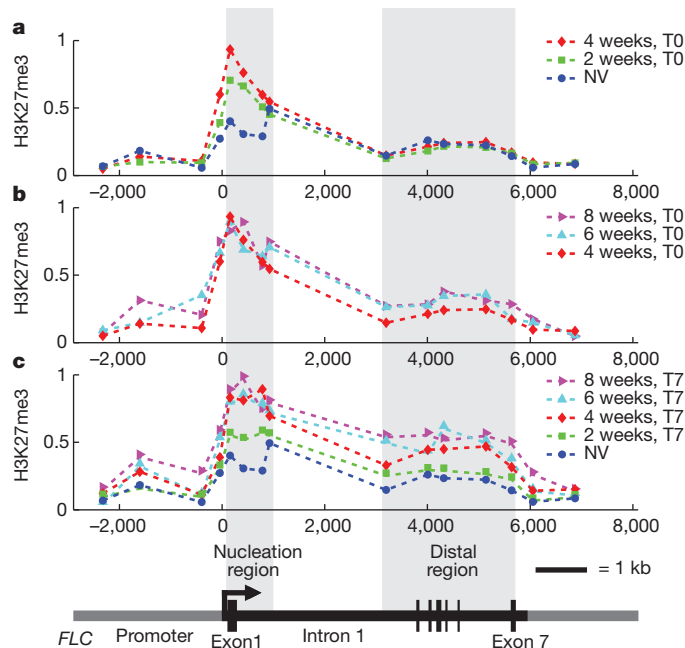
Andrew Angel<sup>1\*</sup>, Jie Song<sup>2\*</sup>, Caroline Dean<sup>2</sup> & Martin Howard<sup>1</sup>

The conserved Polycomb repressive complex 2 (PRC2) generates trimethylation of histone 3 lysine 27 (H3K27me3)<sup>1,2</sup>, a modification associated with stable epigenetic silencing<sup>3,4</sup>. Much is known about PRC2-induced silencing but key questions remain concerning its nucleation and stability. Vernalization, the perception and memory of winter in plants, is a classic epigenetic process that, in *Arabidopsis*, involves PRC2-based silencing of the floral repressor *FLC*<sup>5,6</sup>. The slow dynamics of vernalization, taking place over weeks in the cold, generate a level of stable silencing of *FLC* in the subsequent warm that depends quantitatively on the length of the prior cold. These features make vernalization an ideal experimental system to investigate both the maintenance of epigenetic states and the switching between them. Here, using mathematical modelling, chromatin immunoprecipitation and an *FLC*:GUS reporter assay, we show that the quantitative nature of vernalization is generated by H3K27me3-mediated *FLC* silencing in the warm in a subpopulation of cells whose number depends on the length of the prior cold. During the cold, H3K27me3 levels progressively increase at a tightly localized nucleation region within *FLC*. At the end of the cold, numerical simulations predict that such a nucleation region is capable of switching the bistable epigenetic state of an individual locus, with the probability of overall *FLC* coverage by silencing H3K27me3 marks depending on the length of cold exposure. Thus, the model predicts a bistable pattern of *FLC* gene expression in individual cells, a prediction we verify using the *FLC*:GUS reporter system. Our proposed switching mechanism, involving the local nucleation of an opposing histone modification, is likely to be widely relevant in epigenetic reprogramming.

In *Arabidopsis*, vernalization requires the plant-homeodomain-PRC2 complex (PHD-PRC2), whose components include SWINGER (an E(Z) histone methyltransferase homologue), together with VIN3, VRN5 and VEL1 (PHD proteins required to generate high H3K27me3 levels)<sup>7–10</sup>. PRC2 complexes such as this are thought to initiate trimethylation of H3K27 and then spread this modification through a positive feedback loop where they both bind to H3K27me3 and trimethylate other nucleosomes<sup>7–10</sup>. In principle, this process could generate stable maintenance of histone modifications even when perturbed by DNA replication, where on average half of the modifications could be lost by the insertion of unmodified nucleosomes. In general, however, how to establish stable epigenetic silencing at a level that depends quantitatively on the level of a transient stimulus is not understood.

In an effort to answer this question, we combined chromatin immunoprecipitation (ChIP) experiments with mathematical modelling to dissect the vernalization process. This required high-resolution analysis of H3K27me3 during vernalization as a function of space across the *FLC* locus and time. Previously, we had demonstrated that PRC2 was associated with the whole *FLC* locus before cold, and that this correlated with an epigenetically stable low level of silencing<sup>9,11</sup>. After 8 weeks of cold, a PHD-PRC2 complex had formed at one particular site within the gene and only after plants were returned to the warm did the

PHD proteins associate more generally with PRC2 over the locus. This association resulted in high levels of H3K27me3, which epigenetically maintained the fully silent state. Therefore, we needed to explore the quantitative accumulation of silencing during the cold. Figure 1a (and Supplementary Fig. 1) shows the H3K27me3 profile across *FLC*, in non-vernalized plants and also after 2 and 4 weeks of cold (at 4 °C). A localized peak of H3K27me3 develops (close to exon 1, see Fig. 1), the level of which increases with increasing length of cold, in agreement with previous results<sup>12</sup>. For longer periods of cold (6 and 8 weeks), this nucleation seems to saturate, with some limited general increase in H3K27me3 levels across the rest of the locus (Fig. 1b and Supplementary Fig. 2). During the cold, the rising levels of H3K27me3 in the nucleation region are correlated with an increase in expression of the PHD protein VIN3 (refs 7, 8), whose concentration levels may therefore constitute one of the primary read-outs of the length of cold. However, VIN3 expression returns to low levels immediately (within one day) after return to the warm: the memory of the cold implicit in the VIN3 concentration levels must therefore be converted into a



**Figure 1 | H3K27me3 ChIP experiments.** **a**, H3K27me3 profile across *FLC* for non-vernalized plants (NV) and after 2 or 4 weeks of cold, with 0 days of warm (T0). **b**, H3K27me3 profile across *FLC* after 4, 6 or 8 weeks of cold, with 0 days of warm (T0). **c**, H3K27me3 profile across *FLC* for non-vernalized plants and 7 days post-cold (T7) after 2, 4, 6 or 8 weeks of cold (top) and schematic layout of *FLC* gene (bottom), with definitions of nucleation and distal regions. Panels (a–c) show results from one complete experiment. Similar results were obtained with a biological replicate (Supplementary Figs 1–3).

<sup>1</sup>Department of Computational and Systems Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. <sup>2</sup>Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.

\*These authors contributed equally to this work.

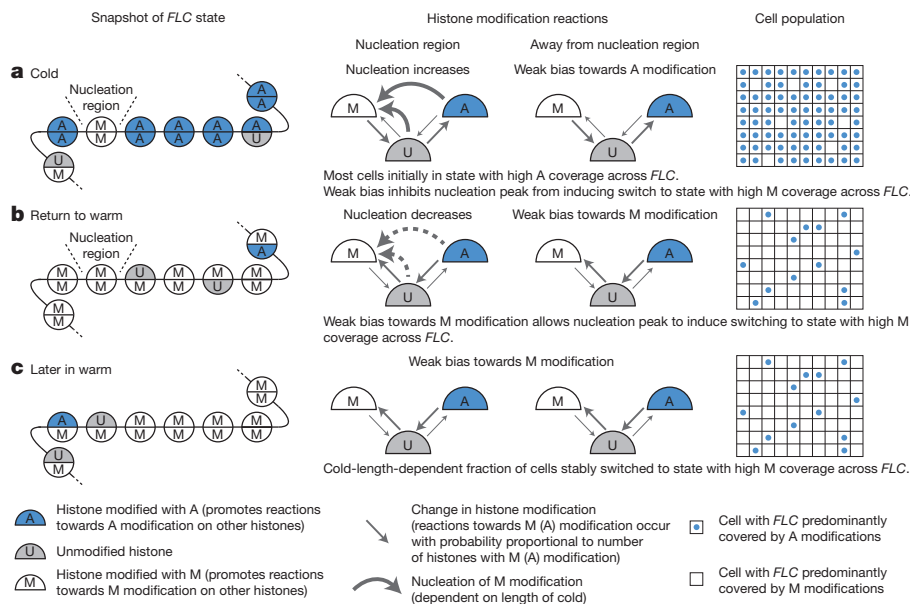


second memory element in the warm, potentially H3K27me3 levels. We measured therefore the H3K27me3 profile across *FLC* 7 days after the plants had been transferred to the warm (Fig. 1c and Supplementary Fig. 3). The profile changed rather little in the nucleation region but rose quantitatively across the rest of the *FLC* locus according to the length of the cold period (Supplementary Fig. 4). This quantitative increase correlated with decreased *FLC* expression (Supplementary Fig. 5), and was consistent with H3K27me3 levels being a key element in epigenetic memory of the cold after return to the warm.

For shorter periods of cold, levels of ChIP-measured H3K27me3 at the nucleation region depended quantitatively on the length of the cold period, and could therefore potentially act to switch quantitatively the epigenetic state of the locus after return to the warm. Epigenetic states are, however, intrinsically stable (in the case of vernalization, over many weeks or months). It was therefore unclear whether the small size of the nucleation region (Fig. 1a) would be sufficient to cause a quantitative switch in the epigenetic state of the *FLC* locus. To answer this question, we turned to mathematical modelling<sup>13–19</sup>. Previous modelling of epigenetic states had focused primarily on cell-autonomous dynamics in yeast<sup>13,15–18</sup>. We wondered whether a similar cell-autonomous mechanism could also explain the quantitative response to the cold seen in vernalization. The above data provided an ideal opportunity to test such a model. In essence, our model is an implementation of a Polycomb-based switch, incorporating highly dynamic rewriting of histone modifications on a timescale of minutes (see Fig. 2 and Supplementary Information for details). The rapid dynamics of histone modifications are vital as they allow for the formation of bistable epigenetic states, despite the noisy turnover of nucleosomes, both continuously (on a timescale of tens of minutes<sup>20</sup>) and discretely at DNA replication (on a timescale of days<sup>21,22</sup>). The model utilizes a simplified set of possible histone modifications: M (H3K27me3), U (unmodified) and A (activating) (other possible assumptions are discussed in the Supplementary Information), where histones with the M or A modification antagonistically promote the transition of other histones towards the same modification, leading to positive feedback<sup>13,16</sup>. At the high levels of noise potentially generated by noisy nucleosome turnover and subsequent random histone mark addition, the presence of spatially long-ranged interactions between histones is an important

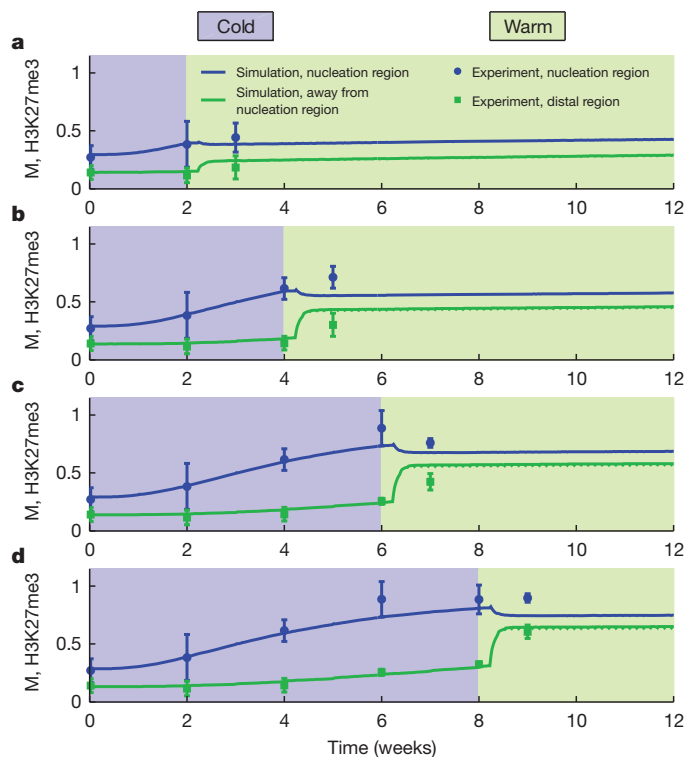
factor in the formation of bistable states<sup>13</sup>. In our model, starting from a state with the *FLC* locus covered by A modifications, two additional processes contribute to the switch-like behaviour at the end of the cold into the opposing state, where the locus is predominantly covered by M modifications. The first process is a site-specific nucleation of the opposing M modification mediated by a cold-induced increase of VIN3 expression (motivated by the localized increase of H3K27me3 levels, see Fig. 1a). The second process is a permanent PHD-PRC2-mediated bias in the histone dynamics towards the M modification on return to the warm, motivated by the observed post-cold spread of the PHD protein VRN5 across the locus<sup>9</sup>. The combination of these two processes enhances the probability of switching between the two bistable epigenetic states, before nucleation is lost (due to clearance of VIN3 on a timescale of days at the end of the cold) and the potential to switch is greatly diminished.

We simulated the above model in 10,000 individual *FLC* loci, and attempted to fit unconstrained model parameters to our experimental ChIP data (see Supplementary Table 2). In Fig. 3a–d we show the simulated population-averaged levels of the M modification averaged separately over the nucleation region (five-histones wide) and the rest of the locus. Also shown are the experimental H3K27me3 values for non-vernalized plants and 0 and 7 days after 2, 4, 6 or 8 weeks of cold, averaged separately over the nucleation and distal regions of *FLC* (as defined in Fig. 1). We find good overall agreement with the observed data. During the cold, some limited increase of H3K27me3 levels does occur in distal regions (Fig. 1b). However, by tracking the dynamics of individual simulated *FLC* loci, we predict that these low levels of H3K27me3 arise from a large subpopulation of cells with little or no H3K27me3 coverage away from the nucleation region and a small subpopulation where H3K27me3 substantially covers the locus (occupying ~90% of histones, limited by noisy dynamics such as nucleosome turnover). Switching into this latter subpopulation is limited by the lack of a (VRN5-dependent) bias towards the M modification. On return to the warm, however, this limitation is removed: our simulations then demonstrate that a transient narrow peak of H3K27me3 can indeed cause the epigenetic state of the locus to be efficiently switched. Again, this rise in H3K27me3 levels is predicted to be due to a subpopulation of cells with substantial coverage of H3K27me3 across the



**Figure 2 | Schematic outline of mathematical model for *FLC* silencing.** a–c, Left, snapshot of *FLC* histone state; middle, histone modification reactions; right, cell population status. a, During the cold, M modifications are added at the nucleation region; most cells are still locked in a state with high A modification coverage. b, At the end of the cold, nucleation dies away, but

histone modification reactions are biased towards M modification. This leads to a short window in which the system can switch from high A to high M modification coverage. c, After the cold, a subpopulation of cells, whose number depends quantitatively on the length of the cold, stably switched to a silenced state, with high levels of M modification coverage.



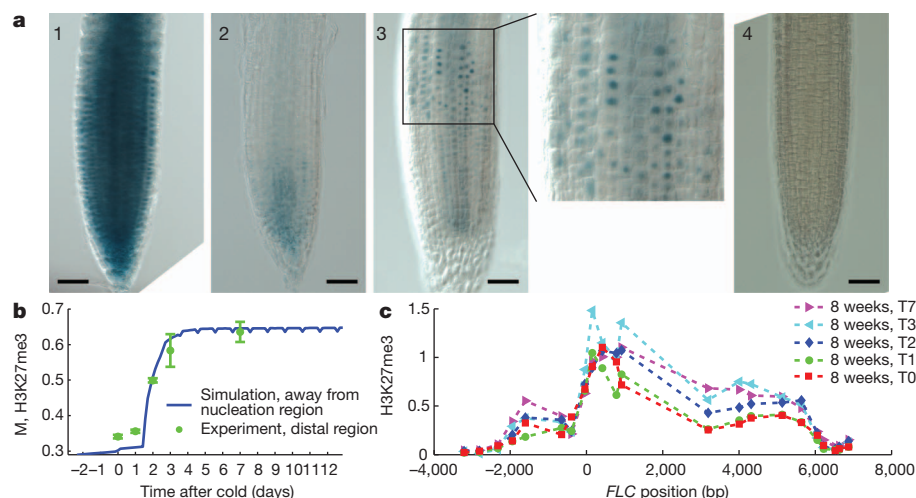
**Figure 3 | Fitting model output to experimental ChIP data.** **a–d**, Simulated population-averaged levels of M modification, averaged separately at and away from the nucleation region, as compared to experimental values of H3K27me3, averaged separately over nucleation and distal regions, for non-vernalized plants and at 0 and 7 days post-cold after 2 (**a**), 4 (**b**), 6 (**c**) or 8 (**d**) weeks of cold. Error bars represent the range,  $n = 2$ .

*FLC* locus, with the remaining cells having low H3K27me3 levels. Simulated population-averaged levels of the M modification from our fitted simulations are found to be approximately stable up to 30 days after the cold (Fig. 3), with a population-level change of less than about 10% from 7 days after the cold, in accordance with the stable long-term silencing observed experimentally. Simulated *FLC* expression levels 7 days after the cold also show a robust reduction from

non-vernalized levels, a reduction whose magnitude increases with increasing length of cold (Supplementary Fig. 5). However, with our fitted parameters, the simulated reduction is not as large as seen experimentally. The model therefore indicates that additional factors could reinforce repression of *FLC*, consistent with a role for the DNA-binding protein VRN1 (ref. 23).

According to the model, quantitative silencing on return to the warm is achieved through the fraction of cells that switch to the silenced state, a fraction that increases with increasing length of cold. Although the active or silenced epigenetic states are stable, a nucleating peak gives rise to a small probability per unit time of switching. This switching is generated by the noisy histone modification dynamics, which are able to amplify the signal from the nucleating peak of H3K27me3 and use a rare fluctuation to flip the epigenetic state of the system. The timescale of a switch at an individual locus from predominant A to M coverage is predicted to be very rapid (about 45 min, see Supplementary Fig. 6).

To test the model prediction of bistable *FLC* expression, we assayed expression of an *FLC:GUS* translational fusion<sup>24,25</sup> in tissue exposed to different cold treatments. For ease of imaging we analysed root tissue, in which cold-induced epigenetic silencing of *FLC* occurs, and which can regenerate into plants that have ‘remembered’ the cold exposure<sup>26</sup>. The results are shown in Fig. 4a and Supplementary Fig. 8, and show that without cold the gene is evenly expressed in the dividing cells of the root tissue. This expression is reduced in all cells after 2 weeks of cold, supporting the fast reduction in *FLC* transcription seen previously<sup>27</sup>. However, on return to the warm, the expression reactivates in a proportion of cells, in a cell-autonomous manner. These cells reflect the proportion of the population where the epigenetic switch to the silent state has not occurred. This heterogeneity in silencing would explain the flowering variation in regenerants from partially vernalized plants<sup>28</sup>. The proportion of cells reactivating *FLC* expression reduces with increasing time in the prior cold (Fig. 4a), fully supporting the model predictions. The diversity of sector patterns reflects the stochastic nature of the silencing process: both single cells and clusters of *FLC:GUS*-expressing cells are observed, the latter probably reflecting clones from a single reactivating cell (Supplementary Fig. 8). The modelling also predicted that population-averaged levels of H3K27me3 could rise rapidly (within 1–2 days) after return to the warm (Fig. 4b). Such increasing levels reflect individual *FLC* loci undergoing very rapid



**Figure 4 | Validating model predictions.** **a**, *Arabidopsis* roots expressing *FLC:GUS* from plants with various cold treatments. Panels from left to right: (1) non-vernalized plants; (2) cold-treated for 2 weeks, harvested immediately on return to warm (T0); (3) cold-treated for 2 weeks, then grown for a further 7 days in the warm (T7); or (4) cold-treated for 4 weeks and then grown for a further 5 days in the warm (T5). Scale bars, 50  $\mu$ m. **b**, Simulated population-averaged levels of M modification, averaged away from nucleation region, as

compared to experimental values of H3K27me3 (from **c** and Supplementary Fig. 9), averaged over distal region 0, 1, 2, 3 or 7 days post-cold after 8 weeks of cold. Error bars represent the range,  $n = 2$ . **c**, Higher time resolution ChIP data for H3K27me3 profile across *FLC* at 0, 1, 2, 3 or 7 days post-cold after 8 weeks of cold. Similar results were obtained with a biological replicate (Supplementary Fig. 9).

switching (taking only about 45 min) into the silenced epigenetic state, but at different starting times. To experimentally test the prediction of a rapid rise at population level, we performed additional higher time resolution (daily) ChIP experiments for the week immediately following the cold (Fig. 4c and Supplementary Fig. 9). In accordance with model predictions, we found a rapid rise of population-averaged H3K27me3 levels, with the H3K27me3 profile increasing mostly between 1 and 3 days after the cold.

Our combined experimental/modelling approach to vernalization has revealed that the quantitative level of epigenetic silencing in the warm is generated by an appropriate proportion of cells switching into a stable *FLC*-silenced state. Controlled switching of individual cells between bistable states, mediated by nucleation of an opposing histone mark may therefore be a common method of achieving a quantitative epigenetic response at a population level.

## METHODS SUMMARY

The ChIP experiments and *FLC* expression analysis were carried out using *Col Fri-sf2* (ref. 29). H3K27me3 and H3 levels were assayed using anti-H3K27me3 (Millipore 07-449) and anti-H3 (Abcam 1791), respectively. *SHOOT MERISTEMLESS (STM)*<sup>30</sup> was used as the internal control for the ChIP experiments (primers are listed in Supplementary Table 1). Data are represented as the ratio of (H3K27me3 *FLC*/H3 *FLC*) to (H3K27me3 *STM*/H3 *STM*). *FLC::GUS* assay was performed using *FLC::CokGUS* in *Ler FRI*<sup>25</sup>. Photoshop adjustment involved only image exposure using adjustment levels.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 January; accepted 26 May 2011.

Published online 24 July 2011.

- Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biol.* **10**, 1291–1300 (2008).
- Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
- Turner, B. M. Defining an epigenetic code. *Nature Cell Biol.* **9**, 2–6 (2007).
- Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nature Rev. Genet.* **11**, 285–296 (2010).
- Sheldon, C. C. *et al.* The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**, 445–458 (1999).
- Michaels, S. D. & Amasino, R. M. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
- Sung, S. & Amasino, R. M. Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* **427**, 159–164 (2004).
- Greb, T. *et al.* The PHD finger protein VRN5 functions in the epigenetic silencing of *Arabidopsis FLC*. *Curr. Biol.* **17**, 73–78 (2007).
- De Lucia, F., Crevillen, P., Jones, A. M., Greb, T. & Dean, C. A. PHD-Polycomb repressive complex 2 triggers the epigenetic silencing of *FLC* during vernalization. *Proc. Natl Acad. Sci. USA* **105**, 16831–16836 (2008).
- Wood, C. C. *et al.* The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl Acad. Sci. USA* **103**, 14631–14636 (2006).
- Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525–535 (2001).
- Finnegan, E. J. & Dennis, E. S. Vernalization-induced trimethylation of histone H3 lysine 27 at *FLC* is not maintained in mitotically quiescent cells. *Curr. Biol.* **17**, 1978–1983 (2007).
- Dodd, I. B., Micheelsen, M. A., Sneppen, K. & Thon, G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129**, 813–822 (2007).
- Salazar, J. D., Foreman, J., Carr, I. A., Rand, D. A. & Millar, A. J. Mathematical model of the epigenetic control of vernalisation in *Arabidopsis thaliana*. *Acta Hort. (ISHS)* **803**, 187–192 (2008).
- Sedighi, M. & Sengupta, A. M. Epigenetic chromatin silencing: bistability and front propagation. *Phys. Biol.* **4**, 246–255 (2007).
- David-Rus, D., Mukhopadhyay, S., Lebowitz, J. L. & Sengupta, A. M. Inheritance of epigenetic chromatin silencing. *J. Theor. Biol.* **258**, 112–120 (2009).
- Mukhopadhyay, S., Nagaraj, V. H. & Sengupta, A. M. Locus dependence in epigenetic chromatin silencing. *Biosystems* **102**, 49–54 (2010).
- Kelemen, J. Z., Ratna, P., Scherrer, S. & Becskei, A. Spatial epigenetic control of mono- and bistable gene expression. *PLoS Biol.* **8**, e1000332 (2010).
- Kaufman, P. D. & Rando, O. J. Chromatin as a potential carrier of heritable information. *Curr. Opin. Cell Biol.* **22**, 284–290 (2010).
- Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
- Reddy, G. V., Heisler, M. G., Ehrhardt, D. W. & Meyerowitz, E. M. Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of *Arabidopsis thaliana*. *Development* **131**, 4225–4237 (2004).
- Grandjean, O. *et al.* In vivo analysis of cell division, cell growth, and differentiation at the shoot apical meristem in *Arabidopsis*. *Plant Cell* **16**, 74–87 (2004).
- Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. *Science* **297**, 243–246 (2002).
- Bastow, R. *et al.* Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427**, 164–167 (2004).
- Sheldon, C. C. *et al.* Resetting of *FLOWERING LOCUS C* expression after epigenetic repression by vernalization. *Proc. Natl Acad. Sci. USA* **105**, 2214–2219 (2008).
- Burn, J. E., Bagnall, D. J., Metzger, J. D., Dennis, E. S. & Peacock, W. J. DNA methylation, vernalization, and the initiation of flowering. *Proc. Natl Acad. Sci. USA* **90**, 287–291 (1993).
- Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **462**, 799–802 (2009).
- Pierik, R. L. M. in *Cellular and Molecular Aspects of Floral Induction* (ed. Bernier, G.) 409–415 (Longmans, 1970).
- Lee, I., Michaels, S. D., Masshardt, A. S. & Amasino, R. M. The late-flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. *Plant J.* **6**, 903–909 (1994).
- Zhang, X. *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank all members of the C.D. and M.H. groups for discussions. We also thank S. Costa for suggestions to improve the *FLC::GUS* imagery and V. Grieneisen, S. Maree, R. Morris, S. Swiezewski and P. Wigge for comments on the manuscript. This research was supported by an Advanced Investigator European Research Council grant and the Core Strategic Grant from the Biotechnology and Biological Sciences Research Council to the John Innes Centre. M.H. also acknowledges support from The Royal Society.

**Author Contributions** C.D. and M.H. conceived the study, A.A., J.S., C.D. and M.H. designed the experiments, J.S. performed the experiments, A.A. and J.S. analysed the experimental data, A.A. and M.H. designed the numerical simulations, A.A. performed the simulations and analysed the simulation data. A.A., J.S., C.D. and M.H. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.D. ([caroline.dean@jic.ac.uk](mailto:caroline.dean@jic.ac.uk)) or M.H. ([martin.howard@jic.ac.uk](mailto:martin.howard@jic.ac.uk)).



## METHODS

**Plant material and growth conditions.** The *Columbia* line *FRI-Sf2* was described previously<sup>29</sup>. Plant growth conditions were described previously<sup>9</sup>. Plants were vernalized for 2, 4, 6 or 8 weeks for the ChIP experiments, together with the non-vernalized control. T0 seedlings were harvested immediately after prolonged cold, whereas T1, T2, T3 and T7 seedlings were grown for 1, 2, 3 and 7 days, respectively, after transfer back to the warm.

**ChIP and real-time quantitative PCR analysis.** ChIP assays were performed as previously described<sup>9</sup> with modifications. Trimethyl-histone H3 lysine 27 was assayed using anti-trimethyl-histone H3 lysine 27 from Millipore/Upstate (catalogue no. 07-449). Histone H3 levels were assayed using anti-H3 core antibody from Abcam (catalogue no. 1791). After immunoprecipitation, DNA was recovered using Chelex 100 resin (Bio-Rad, 10 g per 100 ml ddH<sub>2</sub>O). All ChIP experiments were quantified by quantitative PCR (qPCR) in triplicates with appropriate primers (Supplementary Table 1). *SHOOT MERISTEMLESS* (*STM*)<sup>30</sup> was used as the internal control for the ChIP experiments. Data are represented as the ratio of (H3K27me3 *FLC*/H3 *FLC*) to (H3K27me3 *STM*/H3 *STM*).

**Expression analysis.** cDNA was synthesized using SuperScript III (Invitrogen) with a mixture of oligo d(T) and gene-specific primer (*FLC\_unspliced\_R* 5'-cttgtaatacaagtg gagagc-3'), and analysed by qPCR on a LightCycler 480 II instrument (Roche), using LightCycler 480 SYBR Green mix (Roche). *FLC* sense unspliced (unspliced intron 2/3, *FLC\_unspliced\_F* 5'-cgcaatttcatagcccttg-3', *FLC\_unspliced\_R*) quantification was normalized against the *Arabidopsis* *UBC* gene (*At5g25760*, *UBC\_F* 5'-ctgcgactcagg gaatcttctaa-3', *UBC\_R* 5'-ttgtgccattgaattgaacc-3').

***FLC:GUS* assay.** The *FLC-Col:GUS* in *Ler FRI* was described previously<sup>25</sup>. The *FLC-Col:GUS* translational fusion contains the  $\beta$ -glucuronidase coding sequences cloned into the *NheI* site within *Col FLC* exon 6 (ref. 24). GUS staining was carried out by 4.5 h incubation at 37 °C in X-Gluc staining solution (0.5 mg ml<sup>-1</sup> X-Gluc, 2.0 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 2.0 mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 100 mM phosphate buffer pH 7.0, 0.1% [v:v] Triton X-100, 10 mM EDTA). Seedlings were vacuum infiltrated for 5 min before incubation. After staining, seedlings were cleared in 75% (v:v) ethanol and then 8:1:3 (w:w:w) chloral hydrate:glycerol:water. Roots were imaged using DIC optics on a Leica DM6000 microscope with a Leica DFC420 camera controlled via Leica LAS AF7000 software. Photoshop adjustment involved only image exposure using adjustment levels.

# Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching

Marcelo J. Berardi<sup>1,2</sup>, William M. Shih<sup>2,3,4</sup>, Stephen C. Harrison<sup>1,2,5</sup> & James J. Chou<sup>1,2</sup>

Mitochondrial uncoupling protein 2 (UCP2) is an integral membrane protein in the mitochondrial anion carrier protein family, the members of which facilitate the transport of small molecules across the mitochondrial inner membrane<sup>1,2</sup>. When the mitochondrial respiratory complex pumps protons from the mitochondrial matrix to the intermembrane space, it builds up an electrochemical potential<sup>2</sup>. A fraction of this electrochemical potential is dissipated as heat, in a process involving leakage of protons back to the matrix<sup>2</sup>. This leakage, or ‘uncoupling’ of the proton electrochemical potential, is mediated primarily by uncoupling proteins<sup>2</sup>. However, the mechanism of UCP-mediated proton translocation across the lipid bilayer is unknown. Here we describe a solution-NMR method for structural characterization of UCP2. The method, which overcomes some of the challenges associated with membrane-protein structure determination<sup>3</sup>, combines orientation restraints derived from NMR residual dipolar couplings (RDCs) and semiquantitative distance restraints from paramagnetic relaxation enhancement (PRE) measurements. The local and secondary structures of the protein were determined by piecing together molecular fragments from the Protein Data Bank that best fit experimental RDCs from samples weakly aligned in a DNA nanotube liquid crystal. The RDCs also determine the relative orientation of the secondary structural segments, and the PRE restraints provide their spatial arrangement in the tertiary fold. UCP2 closely resembles the bovine ADP/ATP carrier (the only carrier protein of known structure<sup>4</sup>), but the relative orientations of the helical segments are different, resulting in a wider opening on the matrix side of the inner membrane. Moreover, the nitroxide-labelled GDP binds inside the channel and seems to be closer to transmembrane helices 1–4. We believe that this biophysical approach can be applied to other membrane proteins and, in particular, to other mitochondrial carriers, not only for structure determination but also to characterize various conformational states of these proteins linked to substrate transport.

The elucidation almost 50 years ago of the mechanism of oxidative phosphorylation provided the first example of the coupling of membrane transport to energy production<sup>5</sup>. The mitochondrial respiratory complex transfers electrons from nutrients to oxygen, pumping protons from the mitochondrial matrix to the intermembrane space as it builds up an electrochemical potential of ~200 mV (Supplementary Fig. 1). Most of this potential is converted to work that drives ATP synthesis, but a fraction is dissipated as heat in the uncoupling of the proton electrochemical potential by members of a family of proteins in the mitochondrial inner membrane, collectively termed uncoupling proteins<sup>2</sup> (UCPs). The proton translocation activity of UCPs requires fatty acids as cofactors; it is inhibited by GDP. UCP1 was first identified in the mitochondria of brown adipose tissue as a protein responsible for thermoregulation mediated by such tissue<sup>6</sup>. Unlike UCP1, the primary function of UCP2 may be related to lowering production of reactive oxygen species<sup>2</sup>. Expression of UCP2 in pancreatic  $\beta$ -cells is strongly correlated with decreased insulin secretion, and UCP2

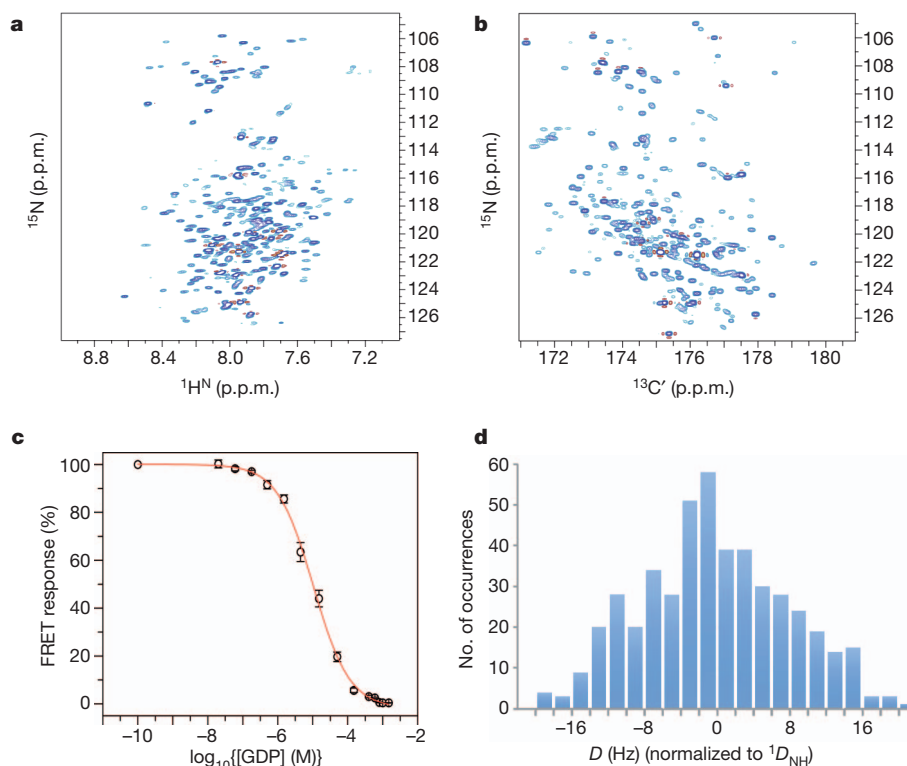
inactivation restores the glucose-sensing pathway<sup>7</sup>. UCP2 also has a role in glucagon secretion from pancreatic  $\alpha$ -cells<sup>8</sup> and in regulation of hunger by means of ghrelin in neurons containing neuropeptide Y and agouti-related protein<sup>9</sup>. More recently, UCP2 has been linked to chemotherapy resistance and metabolic reprogramming in cancer<sup>10,11</sup>.

NMR spectroscopy has yielded the structures of a number of membrane proteins, including channels<sup>12–14</sup>, a redox protein<sup>15</sup>, a kinase<sup>16</sup> and sensory rhodopsin<sup>17</sup>. In many of these studies, a major issue was the purification of the protein and its reconstitution in a model membrane medium supporting protein solubility and stability. In the case of UCP2, reconstitution in dodecylphosphocholine (DPC) micelles with a small amount of lipid (cardiolipin and dimyristoyl phosphatidylcholine (DMPC)) yielded usable NMR spectra (Fig. 1a, b). We prepared homogeneous UCP2 by passing detergent-solubilized protein through a GDP-analogue column, which selects for native protein. Samples used for NMR measurements were at pH 6.5 and contained 5 mM GDP, 150 mM DPC, 1 mM cardiolipin and 2 mM DMPC. Under those conditions, a fluorescence resonance energy transfer (FRET) experiment showed that GDP binds to UCP2 with an apparent dissociation constant of ~5  $\mu$ M (Fig. 1c). When the NMR sample was incorporated into liposomes, it showed fatty-acid-activated and GDP-inhibited proton translocation activity (Supplementary Fig. 2). Negative-stain electron microscopy showed monodisperse, channel-like particles (Supplementary Fig. 3).

Local structure determination by NMR typically involves assigning nuclear Overhauser enhancements between backbone amide protons and aliphatic protons. In the case of UCP2, it was difficult to collect a sufficient number of unambiguous nuclear Overhauser enhancements because of severe overlap of backbone (<sup>1</sup>H, <sup>15</sup>N) resonances. The peaks in the 3D HNCO spectrum are mostly separated, however. The HNCO experiment, which correlates the chemical shifts of backbone <sup>1</sup>H<sup>N</sup>, <sup>15</sup>N and <sup>13</sup>C' nuclides, has been widely used for measuring one-bond RDCs. For structure determination, we therefore relied on orientation restraints derived from RDCs. In a macromolecule marginally oriented in a magnetic field, the RDC between a pair of spin-1/2 nuclides encodes orientations of the internuclear vector connecting the two nuclides. The most effective medium for weakly orienting protein in solution is a liquid crystal formed by large, elongated particles<sup>18</sup>; for membrane proteins, these particles must also resist a high concentration of detergent. We used DNA nanotubes, constructed by ‘DNA origami’, which can form a stable nematic phase unaffected by detergent<sup>19</sup> (Supplementary Fig. 4). The nanotube alignment medium allowed us to measure RDCs for internuclear vectors <sup>1</sup>H<sup>N</sup>–<sup>15</sup>N (<sup>1</sup>D<sub>NH</sub>), <sup>13</sup>C'–<sup>13</sup>C $\alpha$  (<sup>1</sup>D<sub>C'C $\alpha$</sub> ) and <sup>15</sup>N–<sup>13</sup>C' (<sup>1</sup>D<sub>NC'</sub>) (Fig. 1d and Supplementary Fig. 5). On average, there were 2.2 RDCs per residue for regions with confirmed resonance assignment.

With fewer than three RDCs per residue, calculating the local structures using the restrained molecular dynamics protocol presents a severe local minimum problem, because of the intrinsic orientation degeneracy of the dipolar coupling function. We used an approach that

<sup>1</sup>Jack and Eileen Connors Structural Biology Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>4</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02138, USA. <sup>5</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | NMR spectra, GDP binding and outline of RDC-based molecular fragment assignment.** **a**, The  $^1\text{H}$ - $^{15}\text{N}$  transverse-relaxation-optimized spectroscopy (TROSY) heteronuclear single quantum coherence spectrum of  $^{15}\text{N}$ -,  $^{13}\text{C}$ - and  $^2\text{H}$ -labelled UCP2 reconstituted in DPC micelles (with 5 mM GDP), recorded at pH 6.5, 33 °C and a  $^1\text{H}$  frequency of 600 MHz. **b**, The ( $^{15}\text{N}$ ,  $^{13}\text{C}'$ ) projection of the three-dimensional (3D) TROSY-HNCO spectrum of the sample in **a**, recorded under the same conditions. p.p.m., parts per million.

exhaustively searches the Protein Data Bank for fragments of structure that agree with the experimental data. This method, commonly known as molecular fragment replacement (MFR), was first applied in crystallography to build molecular fragments into crystallographically determined electron density<sup>20</sup>. A later NMR study demonstrated that with about four RDCs per residue, it was possible to determine the backbone structure of ubiquitin using molecular fragments fit to RDCs<sup>21</sup>. More recently, a fragment search method that combines experimentally determined chemical shift and RDCs has been implemented in the CS-ROSETTA structure modelling software and shown to be effective in determining structures of water-soluble proteins<sup>22,23</sup>.

We constructed a database containing 320,000 seven-residue fragments extracted from structures of 1,279 water-soluble and 235 membrane-associated proteins. For each seven-residue stretch along the UCP2 sequence, we fitted the corresponding RDCs to all fragments in the database, using singular value decomposition<sup>24</sup>. We evaluated the quality of fit with the parameter  $Q_{\text{free}}$ , which is a normalized root mean squared difference between experimental RDCs and RDCs predicted independently by the structural model<sup>25</sup>. After this exhaustive search, we collected fragments with  $Q_{\text{free}} < 25\%$  and for which the magnitude and rhombicity of the alignment tensor were within 10% of their true values (Fig. 1d). These 'candidate fragments' were used to determine local backbone structure according to a three-step protocol. The first step is 'fragment assignment', whereby the candidate fragments are sorted and those with the smallest  $Q_{\text{free}}$  values are assigned to the corresponding protein segments until no more fragments can be placed without overlap with already assigned segments (Fig. 2a). The second step is 'gap filling'. Here, for any of the gap regions (with fewer than four residues) not assigned in step one, we search for the candidate fragment that overlaps in the backbone dihedral angles  $\phi$  and  $\psi$  with the two flanking fragments and that provides the lowest  $Q_{\text{free}}$

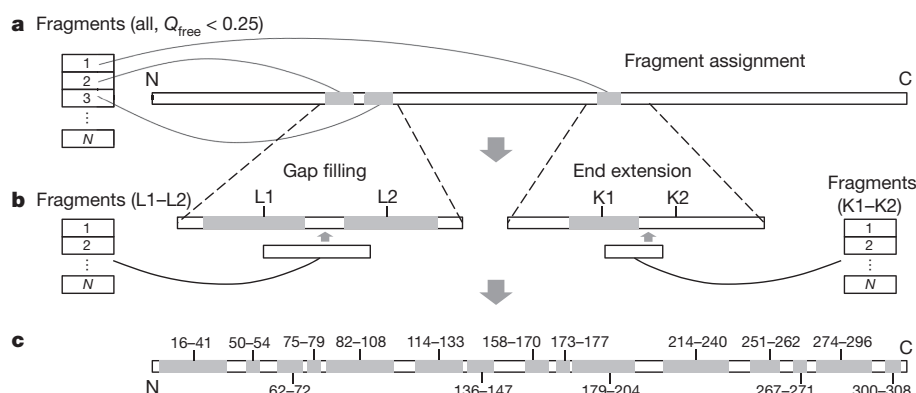
Comparison of **a** and **b** illustrates that most resonances can be resolved in the 3D HNCO spectrum. **c**, Specific binding of GDP to UCP2 under the conditions in **a**. The FRET response (Methods) follows displacement of fluorescent MANT-GDP by GDP as the concentration of the latter increases. [GDP], GDP concentration. **d**, Histogram of 470 unambiguously assigned vectors  $^1D_{\text{NH}}$ ,  $^1D_{\text{C}'\alpha}$  or  $^1D_{\text{NC}'}$  (all normalized to  $^1D_{\text{NH}}$ ). The magnitude and rhombicity of the alignment tensors are 10 Hz and 0.61, respectively.

value for the merged segment consisting of the newly filled gap structure and the two original flanking fragments (Fig. 2b). In the third step, 'end extension', we extend the amino- or carboxy-terminal end of an assigned segment by searching for a fragment that overlaps best with the original segment in  $\phi$  and  $\psi$ , and that provides the lowest  $Q_{\text{free}}$  value for the final extended segment (Fig. 2b). Steps two and three were repeated until no more structure could be assigned with confidence. Using this procedure, we identified 15 structured segments (Fig. 2c).

In addition to RDCs, we obtained semiquantitative distance restraints from PRE measurements. The paramagnetic moiety introduced for PRE measurement is the nitroxide spin label in MTSL (Methods), which we covalently attached at cysteines. PRE restraints have been demonstrated in low-resolution structure determination of both water-soluble and membrane proteins<sup>16,26</sup>. Four UCP2 samples were produced, each with a spin label attached at a unique position (Cys 68, Cys 105, Cys 202 and Cys 255); they all showed similar NMR spectra (Supplementary Fig. 6). These samples together provided 452 PRE distance restraints (an important subset is shown in Fig. 3a).

Finally, to determine the tertiary structure, we strongly enforced the backbone  $\phi$  and  $\psi$  of the 15 MFR-derived structural segments while applying the PRE and RDC restraints. During the structure calculation, the MFR segments made up most of the local structure of the protein, and the PRE and RDC measurements provided spatial and orientational restraints, respectively, for these structured segments. The calculation generated an ensemble of 15 structures with a backbone root mean squared deviation of 1.3 Å (Fig. 3b). There are no experimental data on the side chains. The structure was determined in the presence of GDP, a well-known inhibitor of UCP2 activity. Although side-chain resonance assignments, which are required to obtain an atomic-resolution view of GDP binding, were not available, we investigated GDP binding qualitatively by the PRE method. We



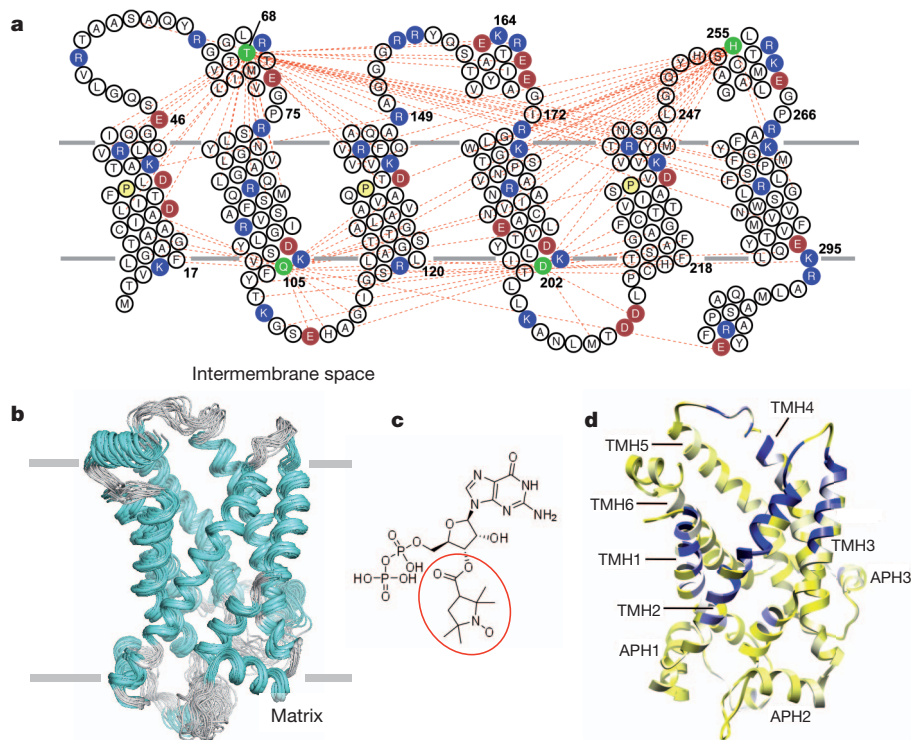


**Figure 2 | Operations involved in RDC-based structural segment building.** **a**, Initial fragment assignment. **b**, Gap filling (left) and end extension (right). **c**, The 15 continuous structured segments of UCP2 determined by RDC-based

MFR (shaded and labelled). Details of these operations are described in the main text and in Methods.

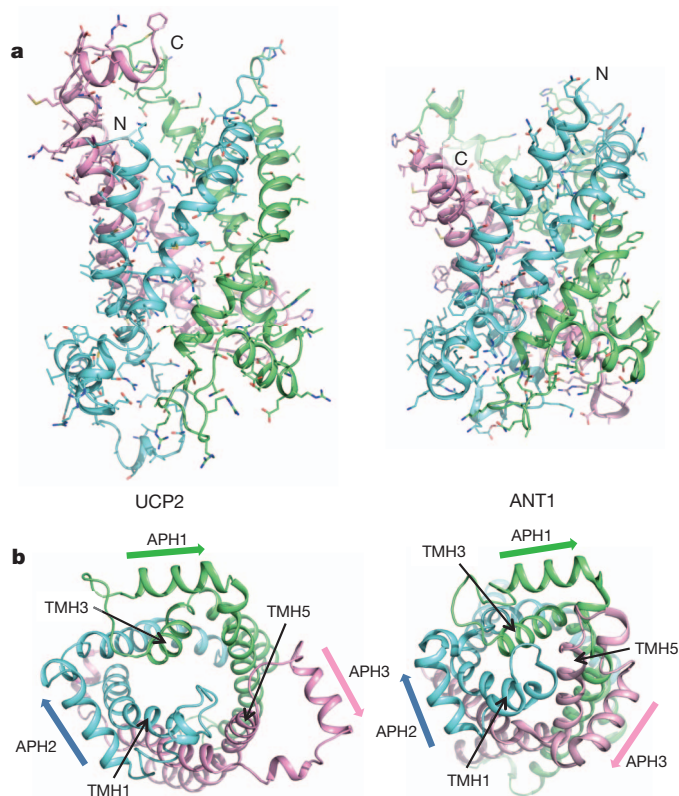
used a paramagnetic nitroxide derivative of GDP (NO-GDP; Fig. 3c) and measured broadening of protein resonances by the spin label. Mapping the magnitude of the PRE to the structure indicates that NO-GDP binds within the UCP2 channel, but the PRE seems to span a large area, covering residues of transmembrane helices (TMHs) 1–4 (Fig. 3d). The distribution of PRE-sensitive residues indicates that the nitroxide radical is in the half of the channel facing the intermembrane space. On the basis of the PRE-derived position of the nitroxide moiety, we located the GDP portion of NO-GDP by maximizing electrostatic interactions between the GDP and charged residues in the channel (Supplementary Fig. 7). The model shows that the PRE data are consistent with GDP binding deep within the UCP2 channel, similar to the proposed ADP-binding site in ANT1<sup>27,28</sup> (also known as SLC25A4).

The overall conformation of UCP2 closely resembles that of the bovine ADP/ATP carrier<sup>4</sup> (ANT1), despite their low (~20%) sequence identity. They are both channel-like structures, in which three pseudo-repeats (Fig. 3a) adopt similar folds (Fig. 4a). Each repeat consists of a TMH (odd numbered), a loop, an amphipathic helix (APH) and a second TMH (even numbered). Moreover, as in ANT1, the TMHs have kinks at prolines that are conserved in the carrier-protein sequences (Fig. 3a). The structural similarity with ANT1 further supports the notion that the members of the large carrier family have a conserved structure and that small variations within the conserved fold govern the specificity of substrate binding and translocation. The most striking differences between the UCP2 and ANT1 structures are in the third repeat. In each of the three repeats of ANT1, the APH packs



**Figure 3 | Solution structure of UCP2 and region of GDP binding.** **a**, UCP2 sequence and membrane topology, with basic and acidic residues shown in blue and red, respectively. The conserved prolines at the proline kinks of TMHs 1, 3 and 5 are shown in yellow. The spin-labelled positions are highlighted in green. The red dashed lines represent long-range or interhelical PRE distances (<19 Å) between the spin-label and backbone amide protons. **b**, An ensemble of 15 low-energy structures derived from NMR restraints. The backbone and

heavy-atom root mean squared deviations for the structured segments in Fig. 2 are 1.2 and 1.8 Å, respectively. **c**, Chemical structure of the spin-labelled GDP, with the paramagnetic nitroxide moiety circled in red. **d**, Mapping the effect of spin-labelled GDP onto the ribbon drawing of UCP2. The colour gradient is from yellow (resonance intensity ratio of broadened to non-broadened,  $\epsilon = 1.0$ ) to white ( $\epsilon = 0.8$ ) to blue ( $\epsilon = 0.3$ ).



**Figure 4 | Comparison of UCP2 and ANT1.** **a**, Side views of UCP2 and ANT1 (Protein Data Bank ID, 1OKC). Pseudo-repeats one (residues 14–112), two (113–210) and three (211–309) are shown in blue, green and pink, respectively (see Fig. 3a for reference). **b**, Views of UCP2 and ANT1 from the matrix side of the carriers, showing loss of threefold pseudo-symmetry in UCP2 as a result of structural differences in repeat three. The orientations of the amphipathic helices are indicated by the arrows.

against the segment of the odd-numbered TMH that follows the conserved proline, and the proline kink in the TMH closes the channel (Fig. 4b). Repeat three of the GDP-bound UCP2 does not follow this pattern: TMH5 seems to have shifted between TMH4 and TMH6 towards the intermembrane side of the carrier (Supplementary Fig. 8). The APH of the same repeat rotates away from the channel by  $\sim 45^\circ$ , and its flanking regions are also substantially different from their counterparts in the other two repeats (Fig. 4b). Consequently, the matrix side of the channel is substantially more open in UCP2 than in ANT1. Although their functional relevance remains to be investigated, the differences we see could have implications for channel opening in proteins of the carrier family.

More than 40 different mitochondrial carriers have so far been identified; they transport a variety of metabolites, nucleotides, ions and vitamins across the inner mitochondrial membrane. Although their overall folds are expected to be similar, obtaining structures of different carriers and different functional states of the same carrier will be important for understanding the mechanism and selectivity of substrate transport<sup>29</sup>. Use of RDCs is well suited to this purpose—for example for describing concerted rearrangements of helices, coupled to opening or closing of the carrier on either side of the membrane—as RDCs are sensitive to the relative orientation of structural segments. We emphasize that new methods of measuring side-chain constraints must be developed to provide higher-resolution NMR structures of these carriers and of membrane proteins of comparable size.

## METHODS SUMMARY

Mouse UCP2 (residues 14–309, with a C-terminal His<sub>6</sub> tag) was expressed using a pET-21 vector in *Escherichia coli* Rosetta DE3 cells. After cell lysis, the lipid

composition of the membrane fraction was adjusted by adding DMPC, cardiolipin and phytanoyl lipids. The protein was extracted using 0.2% DPC in the presence of GDP. The solubilized UCP2 was then subjected to a series of purification steps including Ni-NTA affinity, MonoQ ion exchange, nucleotide-analogue affinity and size exclusion chromatography. The final NMR sample contained 0.8 mM UCP2, 5 mM GDP, 150 mM DPC, 2 mM DMPC, 1 mM cardiolipin, 5 mM  $\beta$ -mercaptoethanol, 30 mM potassium phosphate (pH 6.5), and 80 mM NaCl.

NMR experiments were conducted at 33 °C on spectrometers equipped with cryogenic probes. Sequence-specific assignment of backbone chemical shifts was accomplished using three pairs of triple-resonance experiments and double  $^{15}\text{N}$ -edited nuclear Overhauser enhancement spectroscopy, recorded on protein labelled with  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^2\text{H}$ . RDCs were obtained using an aligned sample containing 0.5 mM UCP2 and 20 mg ml<sup>-1</sup> DNA nanotubes<sup>19</sup> (other components were the same as above).  $^1D_{\text{NH}}$  was measured by *J*-scaled TROSY-HNCO.  $^1D_{\text{C}'\text{C}\alpha}$  and  $^1D_{\text{NC}'}$  were measured by TROSY-HNCO with quantitative modulations of  $J_{\text{C}'\text{C}\alpha}$  (the coupling between backbone C' and C $\alpha$  spins) and  $J_{\text{NC}'}$  (the coupling between backbone N and C' spins), respectively. To obtain PREs, we generated a Cys-less UCP2 mutant and introduced single cysteines at desired positions for labelling with MTSL (Methods). Residue-specific broadening of protein resonances was measured with two TROSY-HNCO spectra, one recorded after nitroxide labelling and another after reducing the nitroxide free electron with ascorbic acid.

Structure determination had two stages: determination of local structural segments using an RDC-based MFR protocol and determination of the spatial arrangement of the MFR-derived segments using PRE distance restraints. Structures were calculated using XPLOR-NIH<sup>30</sup> with backbone  $\phi$  and  $\psi$  values of the assigned structured segments, RDCs and PRE-derived distances. A total of 30 structures were calculated using a simulated annealing protocol, and 15 low-energy structures were selected as the structural ensemble (statistics in Supplementary Table 1).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 December 2010; accepted 1 June 2011.

Published online 24 July 2011.

- Palmieri, F. *et al.* Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochim. Biophys. Acta* **1757**, 1249–1262 (2006).
- Krauss, S., Zhang, C. Y. & Lowell, B. B. The mitochondrial uncoupling-protein homologues. *Nature Rev. Mol. Cell Biol.* **6**, 248–261 (2005).
- Tate, C. G. & Stevens, R. C. Growth and excitement in membrane protein structural biology. *Curr. Opin. Struct. Biol.* **20**, 399–400 (2010).
- Peabay-Peyroula, E. *et al.* Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractylolide. *Nature* **426**, 39–44 (2003).
- Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
- Aquila, H., Link, T. A. & Klingenberg, M. The uncoupling protein from brown fat mitochondria is related to the mitochondrial ADP/ATP carrier. Analysis of sequence homologies and of folding of the protein in the membrane. *EMBO J.* **4**, 2369–2376 (1985).
- Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, beta cell dysfunction, and type 2 diabetes. *Cell* **105**, 745–755 (2001).
- Diao, J. *et al.* UCP2 is highly expressed in pancreatic alpha-cells and influences secretion and survival. *Proc. Natl Acad. Sci. USA* **105**, 12057–12062 (2008).
- Andrews, Z. B. *et al.* UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals. *Nature* **454**, 846–851 (2008).
- Harper, M. E. *et al.* Characterization of a novel metabolic strategy used by drug-resistant tumor cells. *FASEB J.* **16**, 1550–1557 (2002).
- Samudio, I., Fiegl, M. & Andreeff, M. Mitochondrial uncoupling and the Warburg effect: molecular basis for the reprogramming of cancer cell metabolism. *Cancer Res.* **69**, 2163–2166 (2009).
- Schnell, J. R. & Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591–595 (2008).
- Hiller, S. *et al.* Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* **321**, 1206–1210 (2008).
- Wang, J., Pielak, R. M., McClintock, M. A. & Chou, J. J. Solution structure and functional analysis of the influenza B proton channel. *Nature Struct. Mol. Biol.* **16**, 1267–1271 (2009).
- Zhou, Y. *et al.* NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol. Cell* **31**, 896–908 (2008).
- Van Horn, W. D. *et al.* Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* **324**, 1726–1729 (2009).
- Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nature Struct. Mol. Biol.* **17**, 768–774 (2010).
- Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).

19. Douglas, S. M., Chou, J. J. & Shih, W. M. DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. *Proc. Natl Acad. Sci. USA* **104**, 6644–6648 (2007).
20. Jones, T. A. & Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.* **5**, 819–822 (1986).
21. Delaglio, F., Kontaxis, G. & Bax, A. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* **122**, 2142–2143 (2000).
22. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105**, 4685–4690 (2008).
23. Raman, S. *et al.* NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018 (2010).
24. Losonczi, J. A., Andrec, M., Fischer, M. W. F. & Prestegard, J. H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334–342 (1999).
25. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837 (1998).
26. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* **39**, 5355–5365 (2000).
27. Wang, Y. & Tajkhorshid, E. Electrostatic funneling of substrate in mitochondrial inner membrane carriers. *Proc. Natl Acad. Sci. USA* **105**, 9598–9603 (2008).
28. Dehez, F., Pebay-Peyroula, E. & Chipot, C. Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel. *J. Am. Chem. Soc.* **130**, 12725–12733 (2008).
29. Kunji, E. R. & Robinson, A. J. Coupling of proton and substrate translocation in the transport cycle of mitochondrial carriers. *Curr. Opin. Struct. Biol.* **20**, 440–447 (2010).
30. Schwieters, C. D., Kuszewski, J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 66–74 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Oxenoid and R. Sounier for discussions, M. McClintock for help with DNA nanotube preparation, I. Stokes-Rees and P. Sliz for help with computations, and N. Voigt for help with figures. The work was supported by NIH grants 1U54GM094608 (to J.J.C.) and 1DP2OD004641 (to W.M.S.). S.C.H. is an Investigator at the Howard Hughes Medical Institute.

**Author Contributions** M.J.B., W.M.S., S.C.H. and J.J.C. designed the study; M.J.B. prepared NMR samples; M.J.B. and W.M.S. prepared DNA nanotubes; M.J.B. and J.J.C. designed experiments, collected and analysed NMR data, and determined the structure; M.J.B. and J.J.C. wrote the paper; and all authors contributed to editing the manuscript.

**Author Information** The structure of UCP2 has been deposited in the Protein Data Bank under accession number 2LCK. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.J.C. ([chou@cmcd.hms.harvard.edu](mailto:chou@cmcd.hms.harvard.edu)).



## METHODS

**Sample preparation.** Mouse UCP2 (residues 14–309 with a C-terminal His<sub>6</sub> tag) was expressed using a pET-21 vector in *E. coli* Rosetta DE3 cells. After cell lysis, Triton X-100 was added to the lysate at 10  $\mu$ l per gram of cell pellet, followed by stirring at 4 °C for 2 h in the presence of 10  $\mu$ M GDP and 0.2% BME. The suspension was homogenized and centrifuged at 1,500g to remove insoluble debris. To adjust the lipid composition of the membrane fraction, 10 mg DMPC, 2 mg cardiolipin and 0.1 mg phytanoyl lipid per gram of cell pellet were solubilized at 20 mg ml<sup>-1</sup> in 10% octyl glucoside and then added to the supernatant, followed by 2 h of stirring at 4 °C. The membranes were collected by centrifugation at 50,000g for 2 h, and UCP2 was extracted in 40 mM potassium phosphate (pH 8.0), 250 mM NaCl, 50 mM BME, 10  $\mu$ M GDP and 0.2% DPC. The solubilized UCP2 was adsorbed on a Ni-NTA resin in the absence of GDP and eluted with 300 mM imidazole. The sample was then exchanged by dialysis into low-salt buffer and applied to a MonoQ column. The MonoQ flowthrough was incubated with Ciba Blue nucleotide-analogue resin (Sigma) for 1 h, eluted with 100  $\mu$ M GDP and, finally, loaded onto a Sephadex S-200 column for size exclusion purification in 50 mM potassium phosphate (pH 6.5), 100 mM NaCl and 5 mM DPC. The eluted UCP2 sample was supplemented with GDP, detergent and lipids such that the final NMR sample contained 0.8 mM UCP2, 5 mM GDP, 150 mM DPC, 2 mM DMPC, 1 mM cardiolipin, 5 mM BME, 30 mM potassium phosphate (pH 6.5) and 80 mM NaCl.

We have also tested expression and purification of mouse UCP1 and UCP3. UCP1 could be expressed and purified as described for UCP2, and showed FFA-dependent proton translocation activity and GDP inhibition. UCP3 could be expressed but presented serious solubility problems during purification.

**GDP binding.** We used FRET to test GDP binding to UCP2 under the NMR conditions. The intrinsic tryptophan fluorescence ( $\lambda_{\text{ex}} = 280$  nm,  $\lambda_{\text{em}} = 350$  nm) of UCP2 was the resonance energy transfer donor and MANT-GDP ( $\lambda_{\text{ex}} = 356$  nm,  $\lambda_{\text{em}} = 440$  nm) was the acceptor. We recorded the emission intensity at 440 nm ( $\lambda_{\text{ex}} = 280$  nm) of 10  $\mu$ M UCP2/MANT-GDP complex for increasing GDP concentrations. Figure 1c shows the FRET response as a percentage relative to 1:0 MANT-GDP:GDP. Data were fitted to the standard equilibrium binding equation.

**NMR spectroscopy.** NMR experiments were conducted at 33 °C on spectrometers equipped with cryogenic TXI probes (Bruker). Sequence-specific assignment of backbone <sup>1</sup>H<sup>N</sup>, <sup>15</sup>N, <sup>13</sup>C $\alpha$ , <sup>13</sup>C $\beta$ , and <sup>13</sup>C' chemical shifts were accomplished using three pairs of triple-resonance experiments and a double <sup>15</sup>N-edited NOESY, recorded on a protein sample labelled with <sup>15</sup>N, <sup>13</sup>C and <sup>2</sup>H. The triple-resonance experiments include the TROSY versions of HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO and HNCO<sup>31,32</sup>. The <sup>15</sup>N-edited NOESY is a high-resolution, 3D (H<sup>N</sup>, H<sup>N</sup>)-HMQC-NOESY-TROSY with <sup>15</sup>N, <sup>15</sup>N and <sup>1</sup>H<sup>N</sup> evolution in the *t*<sub>1</sub>, *t*<sub>2</sub> and *t*<sub>3</sub> dimensions, respectively.

The RDCs were obtained by subtracting *J* of the isotropic sample above from *J* + *D* of the aligned sample containing 0.5 mM protein and 20 mg ml<sup>-1</sup> DNA nanotubes (all other components were the same as in the isotropic sample). Details of preparing DNA nanotubes and aligned protein samples are in Supplementary Information. The sign of the RDC follows the convention that  $|J_{\text{NH}} + {}^1D_{\text{NH}}| < 92$  Hz when  ${}^1D_{\text{NH}}$  is positive. The <sup>1</sup>H–<sup>15</sup>N couplings were measured at 600 MHz (<sup>1</sup>H frequency) using the *J*-scaled TROSY-HNCO experiment to exploit the favourable relaxation property of the TROSY transitions<sup>33</sup>. In this experiment, two interleaved spectra were recorded, the regular TROSY-HNCO spectrum and a modified TROSY-HNCO spectrum with *J*<sub>NH</sub> evolution during the <sup>15</sup>N chemical shift evolution scaled to zero. The <sup>13</sup>C'–<sup>13</sup>C $\alpha$  couplings were measured at 600 MHz using the 3D TROSY-HNCO experiment with quantitative *J*<sub>C'C $\alpha$</sub>  modulations of 0 and 28 ms (ref. 34). The <sup>13</sup>C'–<sup>15</sup>N couplings were measured at 750 MHz using the 3D TROSY-HNCO experiment with quantitative *J*<sub>NH</sub> modulations of 33 and 66 ms (ref. 35).

To introduce a single paramagnetic site for PRE measurement, the five cysteines of UCP2 were all mutated to alanine or serine (Cys 25 Ala, Cys 191 Ser, Cys 217 Ser, Cys 227 Ser, Cys 256 Ser); the cysteine-free UCP2 had GDP binding properties similar to those of wild-type protein. A single cysteine was introduced into the protein on the basis of the known secondary structures from MFR segments. The

nitroxide bearing compound, *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL), was then attached at the cysteine position by adding sevenfold excess label to 10  $\mu$ M UCP2 in the NMR buffer at pH 8.0, and incubating at 25 °C for 4 h. Excess label was removed to avoid nonspecific broadening. The pH was changed back to 6.5 for NMR measurements. To quantify residue-specific broadening of backbone <sup>1</sup>H<sup>N</sup>, two TROSY-HNCO spectra were recorded, one after nitroxide labelling and the other after reduction of the nitroxide free electron with a  $\times 5$  molar ascorbic acid. The same protocol was used to measure broadening of UCP2 resonances by the spin-labelled GDP (Fig. 3c; custom synthesized by Shanghai ChemPartner).

**Structure determination.** The procedure had two stages: determination of local structural segments by RDC-based MFR and determination of the spatial arrangement of the MFR-derived segments in the tertiary fold using PRE distance restraints. In stage one, we fitted RDCs of every seven-residue stretch along the UCP2 sequence to the fragments in the database using the program PALES<sup>36</sup>. Only those seven-residue stretches with on average more than two RDCs per residue were used for fragment searching. In the end, 5,520 fragments with *Q*<sub>free</sub> < 0.25 were collected. For each candidate fragment (residue *n* – *m*) evaluated in the fragment assignment, gap filling or end extension operation (described in the main text), only the  $\phi$  and  $\psi$  angles of residue (*n* + 1) – (*m* – 1) were used, because the end residues do not have the <sup>15</sup>N–<sup>13</sup>C' dipolar coupling. During fragment assignment, we found that for each seven-residue stretch, the best 5–10 fragments typically have very similar *Q*<sub>free</sub> values (differences less than 10%). Therefore, the best fragment was assigned to the region only if these fragments have a backbone root mean squared deviation of less than <0.5 Å. Similarly, there are multiple fragments that score equally well in gap filling or end extension. In those cases, we assigned the best-scoring fragment only if the final merged or extended segments have backbone root mean squared deviation of less than <1 Å. Examples of fragment convergence are shown in Supplementary Fig. 9.

In stage two, we calculated structures using XPLOR-NIH<sup>37</sup> with the assigned structured segments, RDCs and PREs. The  $\phi$  and  $\psi$  values of the segments in Fig. 2c were strongly enforced by a harmonic potential with force constant ramped from 10 to 1,000 kcal mol<sup>-1</sup> rad<sup>-2</sup>. All RDCs used for determining the segments were applied, and the RDC force constant was ramped from 0.01 to 1.5 kcal mol<sup>-1</sup> Hz<sup>-2</sup>. PRE restraints (uncertainties in Supplementary Table 2) were enforced with flat-well harmonic potentials, with the force constant ramped from 1 to 40 kcal mol<sup>-1</sup> Å<sup>-2</sup>. In addition to experimental restraints, a weak database-derived 'Rama' potential function<sup>38</sup> was ramped from 0.02 to 0.2 (dimensionless force constant) for the general treatment of side-chain rotamers. A total of 30 monomer structures were calculated using a simulated annealing protocol in which the bath temperature was cooled from 2,000 to 200 K. Fifteen low-energy structures were selected as the structural ensemble (Supplementary Table 1).

- Salzmann, M., Wider, G., Pervushin, K. & Wuthrich, K. Improved sensitivity and coherence selection for [N-15,H-1]-TROSY elements in triple resonance experiments. *J. Biomol. NMR* **15**, 181–184 (1999).
- Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).
- Kontaxis, G., Clore, G. & Bax, A. Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J. Magn. Reson.* **143**, 184–196 (2000).
- Jaroniec, C. P., Ulmer, T. S. & Bax, A. Quantitative J correlation methods for the accurate measurement of <sup>13</sup>C'–<sup>13</sup>C $\alpha$  dipolar couplings in proteins. *J. Biomol. NMR* **30**, 181–194 (2004).
- Chou, J. J., Delaglio, F. & Bax, A. Measurement of one-bond <sup>15</sup>N–<sup>13</sup>C' dipolar couplings in medium sized proteins. *J. Biomol. NMR* **18**, 101–105 (2000).
- Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122**, 3791–3792 (2000).
- Schwieters, C. D., Kuszewski, J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 66–74 (2002).
- Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. *J. Magn. Reson.* **125**, 171–177 (1997).

## ADDENDUM

doi:10.1038/nature10219

## A strong ferroelectric ferromagnet created by means of spin–lattice coupling

June Hyuk Lee, Lei Fang, Eftihia Vlahos, Xianglin Ke, Young Woo Jung, Lena Fitting Kourkoutis, Jong-Woo Kim, Philip J. Ryan, Tassilo Heeg, Martin Roeckerath, Veronica Goian, Margitta Bernhagen, Reinhard Uecker, P. Chris Hammel, Karin M. Rabe, Stanislav Kamba, Jürgen Schubert, John W. Freeland, David A. Muller, Craig J. Fennie, Peter Schiffer, Venkatraman Gopalan, Ezekiel Johnston-Halperin & Darrell G. Schlom

*Nature* **466**, 954–958 (2010)

This Letter determined that  $\text{EuTiO}_3$ , when appropriately strained, becomes a strong ferroelectric ferromagnet, in agreement with prediction. Strong ferroelectrics are proper ferroelectrics, having polarization as their order parameter, with high paraelectric-to-ferroelectric transition temperatures ( $T_c$ ). Such ferroelectrics are manifested by a high  $T_c$  and a high peak at  $T_c$  in the dielectric constant versus temperature behaviour, signifying that ferroelectricity is driven by the soft mode, which is indicative of proper ferroelectricity. Our measurements of strained  $\text{EuTiO}_3$  demonstrate both of these characteristics (shown in Fig. 3 of our Letter), and led us to conclude that strained  $\text{EuTiO}_3$  is a strong ferroelectric. In contrast, all well-established prior single-phase ferroelectric ferromagnets are improper or pseudoproper ferroelectrics (that is, with weak ferroelectricity resulting in minuscule  $P_s$ ). We did not present  $P_s$  values in our Letter. Second harmonic generation measurements do not provide quantitative values of  $P_s$  and attempts to determine  $P_s$  via pyroelectric measurements (Yan, L., Li, J. F. & Viehland, D., personal communication)<sup>1</sup> resulted in unphysically high values, presumably owing to electrical leakage. Nonetheless, the magnitude of the  $P_s$  of our strained  $\text{EuTiO}_3$  films can be estimated as follows. In their classic work, Abrahams, Kurtz, and Jamieson<sup>1</sup> established a correlation between  $P_s$  and  $T_c$  for displacive ferroelectrics. By studying numerous displacive ferroelectrics they found<sup>1</sup>

$$T_c = (2.00 \pm 0.09) \times 10^4 (\Delta z)^2 \quad (\text{equation (1) of ref. 1})$$

and

$$P_s = (258 \pm 9) \Delta z \quad (\text{equation (5) of ref. 5})$$

where  $T_c$  is the paraelectric-to-ferroelectric transition temperature in K,  $\Delta z$  is the atomic displacement of the ‘homopolar’ metal atom in Å, and  $P_s$  is the spontaneous polarization of the ferroelectric in  $\mu\text{C cm}^{-2}$ . Combining these equations to eliminate  $\Delta z$  allows  $P_s$  to be estimated from  $T_c$  in displacive ferroelectrics. The huge anomaly of the soft optical phonon near  $T_c$  that we observe (Supplementary Fig. 1 of our Letter) shows that strained  $\text{EuTiO}_3$  is a displacive ferroelectric, making the aforementioned correlation applicable. Plugging in our measured value of  $T_c$  (Fig. 3 in our Letter) yields  $P_s = 29 \pm 2 \mu\text{C cm}^{-2}$  for our strained  $\text{EuTiO}_3$  films from this established correlation. This rough estimate is consistent with our first-principles theoretical predictions— $P_s = 21 \mu\text{C cm}^{-2}$  for  $\text{EuTiO}_3$  under +1.1% biaxial tension, corresponding to the strain of our commensurate  $\text{EuTiO}_3$  films grown on (110)  $\text{DyScO}_3$ . Thus, the data in our Letter shows that appropriately strained  $\text{EuTiO}_3$  is a strong ferroelectric ferromagnet.

1. Abrahams, S. C., Kurtz, S. K. & Jamieson, P. B. Atomic displacement relationship to Curie temperature and spontaneous polarization in displacive ferroelectrics. *Phys. Rev.* **172**, 551–553 (1968).



# CAREERS

**COLUMN** How scientists can sell themselves effectively **p.117**

**JOURNAL ENTRIES** Read about the travails of young scientists [go.nature.com/3fttcj](http://go.nature.com/3fttcj)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

M. J. DORF/CORBIS



INTERDISCIPLINARY STUDIES

## Seeking the right toolkit

*Many researchers are exploring the boundaries of established disciplines. But does this cutting-edge approach lead to job opportunities?*

BY BRYN NELSON

When David Camarillo begins his tenure-track job in the engineering department at Stanford University in California later this year, it will represent not only a career milestone, but a validation of some unconventional training.

Camarillo enrolled as a mechanical-engineering PhD student at Stanford, but an interest in medicine led him to train in a medical-robotics lab affiliated with both the surgery and computer-science departments. He benefited from a fellowship awarded by Stanford's Bio-X programme, which encourages interdisciplinary research and education. Camarillo also did a year-long internship at a start-up

biotechnology company midway through his schooling, and joined the company after graduating, before returning to academia.

Camarillo concedes that an interdisciplinary training route isn't for everyone. "In my case it has worked out OK so far," he says. "I've seen some people do the interdisciplinary path and then go back to their bread and butter."

No one can blame researchers for retreating to a more conventional scheme. Interdisciplinary research institutes are springing up around the world, giving graduate students and postdocs many opportunities for cutting-edge research and cross-fertilization, but employers often prefer to hire scientists who have stayed in established research disciplines.

Barry Bozeman, a policy analyst at the

University of Georgia in Athens who studies scientists' career trajectories, says that for now, an interdisciplinary background is "very rarely an advantage" when looking for a faculty position. Biotechnology and pharmaceutical firms might be more accommodating, as long as the applicant's unconventional research fits within the company's overall scientific aims. But formal interdisciplinary training may be less important than informal learning experiences in labs, institutes and universities that encourage the intermingling of a broad range of ideas.

### MIXED MESSAGES

Interdisciplinary graduate programmes have been around for decades. But until Bio-X began awarding fellowships in 2004, most ►



► programmes did not attempt to integrate the biological and physical sciences, and many served mainly to funnel undecided students toward specific departments.

This year, a white paper co-authored by researchers at the Massachusetts Institute of Technology (MIT) in Cambridge called for a more ambitious fusion of disciplines, seeking to combine technologies and know-how into a powerful model of innovation (P. A. Sharp *et al.* *The Third Revolution: The Convergence of the Life Sciences, Physical Sciences and Engineering* MIT, 2011). That ideal has been embraced by a host of research centres that house researchers from a mixture of fields in shared facilities that promote interaction.

Such institutions include the Clark Center at Stanford, home to the Bio-X programme. The centre's open-plan building hosts 44 labs spanning the gamut of disciplines, although faculty members housed there retain their departmental appointments. "Think of it like Noah's ark. The idea was to put two of everything in there, and let them breed and see what happens," says Carla Shatz, director of Bio-X.

Bozeman says that boundary stretching is most welcome in research fields that have seen a lot of change in recent years, such as biology and engineering. Other fields that might accept an interdisciplinary mindset, such as mathematics, are in relatively low demand by employers. The chances of making practical use of an interdisciplinary degree may be best, says Bozeman, when that research focus evolves into an established field of its own. Booms in synthetic biology, biophysics, nanobiology and systems biology over the past few years suggest that being in the right place at the right time can pay off for a nascent career.

Some universities are trying out hiring initiatives that will benefit unconventional researchers. For example, Michigan Technological University in Houghton has based some of its recruitment decisions over the past five years not on individual departments, but on multidisciplinary research themes. The first of these, sustainability, brought in seven faculty members with nontraditional joint appointments such as social sciences paired with forest resources and environmental science. "They will be the catalysts of increased interactions between the departments," says Max Seel, the university's

provost and head of the hiring initiative. Other themes include computing innovation, energy and health.

Interdisciplinary graduate-training programmes are less common. The few that have been established include Bio-X, the Integrative Systems Biology Doctoral Training Centre (DTC) at the Interdisciplinary Biocentre (MIB) of the University of Manchester, UK, and the Biodesign Institute at Arizona State University in Tempe.

Bio-X fellowships are given to students who have already been accepted into specific departments at Stanford; they receive a training grant but stay in their adviser's home department. "We're not really taking them off their route, but what we're doing is training them to think more broadly," says Heideh Fattaey, executive director of operations and programmes for Bio-X.

The approach has paid off, says Fattaey. Camarillo has been academically successful; and one former Bio-X fellow who worked with an ophthalmologist and a chemical engineer to develop an artificial cornea has since started his own biotech company. A current fellow is working with a mechanical engineer and a physicist on a new method for measuring muscle motion.

Manchester's interdisciplinary training programme has been running only since 2006. Even so, says Nigel Scrutton, director of the MIB, graduates are highly employable and have found jobs in small spin-off biotech firms and large multinational companies alike. And despite the tough economic climate, he says, graduates have excelled in securing externally funded academic fellowships. Many have been hired at Manchester, particularly at the MIB, where faculty members are in a prime position to spot emerging talent.

Most PhDs in Britain take three years, but one at the systems biology DTC takes four: the first year is spent learning subjects ranging from the dynamics of nonlinear systems to eukaryotic transcription and translation. For Ben Small, a PhD student at the DTC, committing to an additional year has been worth it. "It's given me a unique perspective on biological problems and the way that you can approach them," he says. "I never hesitated or had any doubt that this would be the right route."

After his undergraduate degree, Small conducted preclinical and clinical drug-discovery research in the United Kingdom at pharmaceutical companies Eli Lilly in Windlesham and AstraZeneca at Alderley Park. He saw a need to use the tools and techniques of the predictive sciences in pharmacological research, so he enrolled in the DTC and has joined a neurosystems lab to focus on the chemical and computational biology of inflammation.

Small is set to receive his PhD in 2012. After that, he hopes to start a research career in either industry or academia, focusing on systems-biology-based pharmacology. For now, his first



**"If people think that what you're proposing to do is vigorous and interesting, then you will get the job."**

Polina Anikeeva

Polina Anikeeva, newly hired as a materials scientist at MIT, plays down the benefits of an interdisciplinary background for securing a faculty post. "If you have ideas that the department likes and people think that what you're proposing to do is vigorous and interesting, then you will get a job," says Anikeeva, who did her postdoctoral research at the Clark Center. "I don't think it really depends on if you have interdisciplinary training or not." But she acknowledges that she was drawn to her Bio-X-affiliated lab by the opportunity to break into the medical-devices arena without previous biology training. At the Clark Center, Anikeeva built implantable devices that can measure electronic signals from individual neurons. Last year, she landed two tenure-track job offers, one from MIT and another from the Swiss Federal Institute of Technology in Lausanne, to expand her work on designing implants and prosthetics.

## HOME BASE

Bozeman cautions that although it is relatively easy for universities to hire people with interdisciplinary backgrounds for postdocs, it is much harder to get interdisciplinary faculty positions. That could lead scientists without a history of close affiliation with an established department to serial fellowships and postdoc limbo, or job-hunting challenges in the broader market.

"If they're in an interdisciplinary programme, there may be concerns about whether they can teach at the undergraduate level, whether their research is really in a specific area that's relevant to that [hiring] department," says Jonathan Dordick, director of the Center for Biotechnology and Interdisciplinary Studies at Rensselaer Polytechnic Institute in Troy, New York. "Falling through the cracks may in fact be fairly common, and it may be more difficult for people like that to be able to get a faculty position."

Recognizing such difficulties, the Interdisciplinary Studies Graduate Program at the University of British Columbia in Vancouver,



**"People who establish interdisciplinary degrees are also more likely to hire people with interdisciplinary degrees."**

Barry Bozeman

Canada, warns students that they must have a home department. This “gives them full citizenry in terms of access to financial and physical resources”, says the programme’s website. And when they complete their graduate studies, students are “strongly advised to be strategic about their post-doctoral placement, since most must find a job in an existing more traditional field”.

### FRINGE BENEFITS

Despite the caveats, pursuing a course that defies traditional boundaries can still yield valuable, if subtle, benefits. Sam Hay, a biophysicist at the University of Manchester, wasn’t intent on interdisciplinary studies when he sought a postdoctoral position five years ago, but the MIB was one of the only places hosting research that really interested him: studying the quantum mechanics of biological reactions. “There was a lab set up and running, doing what I wanted to do,” says Hay. But “the best things about it are the sort of things you don’t really expect”, he adds.

Those unexpected benefits include communal spaces and seminar programming designed to foster the creative intermingling of ideas. “You walk into a seminar or lecture and realize it’s somebody working on a technique that’s a hell of a lot better than the one you’ve been trying to wrap your data around,” says Hay. That atmosphere, he says, contributed to the success of his postdoc. Last September, he received a coveted academic fellowship at the biocentre, an entry route to the equivalent of a tenure-track research position. “There are quite a few people floating around here that don’t obviously fit in any one department,” says Hay. “Ultimately, time will tell whether this is a fad, but I suspect it probably won’t be.”

Fattaey laughs when asked whether the concept of interdisciplinary training is gaining in popularity. “Today I have already talked to three groups that want to do this on a daily basis,” she says, citing delegations from South Korea, Belgium and Denmark. She also lists universities or consortia in five US states that have requested her advice on launching programmes similar to Bio-X.

Even sceptics concede that the increase in popularity could pay dividends for graduate students and postdocs. “People who establish interdisciplinary degrees are also more likely to hire people with interdisciplinary degrees,” says Bozeman. Fattaey foresees a thriving network of like-minded universities and researchers, further adding to the talent pool and job prospects. “We’re not going anywhere if all of us don’t work together,” she says. ■

**Bryn Nelson** is a freelance writer based in Seattle, Washington.

## COLUMN

# Scientists for sale

There are ways for researchers to sell themselves — but they shouldn’t overdo it, says **Mariano A. Loza-Coll**.

**I**t all started well. At an interview for a faculty position, a colleague of mine provided an account of his work as a postdoc. Afterwards, a member of the recruiting committee praised him and provided feedback on his presentation. Then he said it: “Great job, but you should sell yourself better.”

When I was growing up in Argentina, before the advent of online commerce, people would place empty cans on the roofs of their cars to indicate that a vehicle was for sale. When I hear stories like my friend’s, I chuckle at the thought of showing up to job interviews with an empty plum-tomato tin on my head.

I’ve heard the sales argument before. As scientists, we need to ‘sell our work’ (or, my favourite, we need to ‘make it sexy’). I get what that means; we need to highlight the most salient findings and implications of our work without hiding its caveats and limitations. Instead, we should turn them into mere blemishes under the flattering light of our discoveries.

But how should we go about that? Or, more importantly, how can we avoid overdoing it? When I hear that I need to sell my work, my grant proposal and even myself, I can’t shake the fear that if I push too hard, I will start sounding like the proverbial used-car salesman.

I decided to approach the question as any sensible scientist would: I googled ‘key steps to a successful sale’. A few tips caught my eye, because scientists tend to overlook them. They can be used when applying for a faculty position or research grant, or when pitching a paper to an academic journal.

First, be realistic and make sure that your product fits the needs of your target audience. Sales associates understand that they will not be able to sell everything to everyone. I don’t only mean trim the sub-par science; a tough sale won’t always be overcome by polishing your product. Even your most elegant science may not fly with an audience ‘not in the market’ for it, so be sure to pick your target wisely.

Second, a sales meeting is a conversation. All the tips I found stressed that the salesperson must listen to potential buyers to understand

their needs. It might seem obvious, but it got me thinking about how often we scientists make the process mostly about ourselves (my CV, my publications, my recommendation letters, me, me, me...). Some experts even advise forcing the conversation if it isn’t part of the interview or application process.

For example, they might arrange a series of phone interviews to get to know their ‘buyers’ — the hiring committee, the grant makers, the journal editors — before the interview or submission. But remember that looking up faculty members online doesn’t count as a conversation.

Finally, explain clearly what will happen after the sale. Buyers need to know how they will put you, the product, to use. Think of yourself as a new printer. Are you ‘upgradable’? Your prospective employers might want to know how easily you can scale your lab up or down, or move it between floors or buildings if necessary. They may also appreciate knowing that you’re ready to lecture on several subjects at a moment’s notice. Do you come with a ‘service contract’? Try to demonstrate that you can count on a network of collaborators in case of problems. What are your ‘consumables’? If your research involves the use of unconventional materials, make sure that you show that you have thought carefully about how to secure them. Do you come with ‘pre-installed drivers’? Be honest about what you need to get started. It’s best to tell your department about the particle accelerator you’ll need in your basement before the fleet of moving trucks arrives.

These universal sales principles won’t apply to every case, but they could come in handy, especially in a tough job market or funding climate. Of course, should everything else fail, you can always break out the car-salesman routine. Look the search-committee members squarely in the eye, give them your widest grin and ask, “Say, what will it take for me to get this job today?” ■

**Mariano A. Loza-Coll** is a postdoc in genetics at the Salk Institute for Biological Studies in La Jolla, California.



P. GILLIGAN



Canada, warns students that they must have a home department. This “gives them full citizenry in terms of access to financial and physical resources”, says the programme’s website. And when they complete their graduate studies, students are “strongly advised to be strategic about their post-doctoral placement, since most must find a job in an existing more traditional field”.

### FRINGE BENEFITS

Despite the caveats, pursuing a course that defies traditional boundaries can still yield valuable, if subtle, benefits. Sam Hay, a biophysicist at the University of Manchester, wasn’t intent on interdisciplinary studies when he sought a postdoctoral position five years ago, but the MIB was one of the only places hosting research that really interested him: studying the quantum mechanics of biological reactions. “There was a lab set up and running, doing what I wanted to do,” says Hay. But “the best things about it are the sort of things you don’t really expect”, he adds.

Those unexpected benefits include communal spaces and seminar programming designed to foster the creative intermingling of ideas. “You walk into a seminar or lecture and realize it’s somebody working on a technique that’s a hell of a lot better than the one you’ve been trying to wrap your data around,” says Hay. That atmosphere, he says, contributed to the success of his postdoc. Last September, he received a coveted academic fellowship at the biocentre, an entry route to the equivalent of a tenure-track research position. “There are quite a few people floating around here that don’t obviously fit in any one department,” says Hay. “Ultimately, time will tell whether this is a fad, but I suspect it probably won’t be.”

Fattaey laughs when asked whether the concept of interdisciplinary training is gaining in popularity. “Today I have already talked to three groups that want to do this on a daily basis,” she says, citing delegations from South Korea, Belgium and Denmark. She also lists universities or consortia in five US states that have requested her advice on launching programmes similar to Bio-X.

Even sceptics concede that the increase in popularity could pay dividends for graduate students and postdocs. “People who establish interdisciplinary degrees are also more likely to hire people with interdisciplinary degrees,” says Bozeman. Fattaey foresees a thriving network of like-minded universities and researchers, further adding to the talent pool and job prospects. “We’re not going anywhere if all of us don’t work together,” she says. ■

**Bryn Nelson** is a freelance writer based in Seattle, Washington.

## COLUMN

# Scientists for sale

There are ways for researchers to sell themselves — but they shouldn’t overdo it, says **Mariano A. Loza-Coll**.

**I**t all started well. At an interview for a faculty position, a colleague of mine provided an account of his work as a postdoc. Afterwards, a member of the recruiting committee praised him and provided feedback on his presentation. Then he said it: “Great job, but you should sell yourself better.”

When I was growing up in Argentina, before the advent of online commerce, people would place empty cans on the roofs of their cars to indicate that a vehicle was for sale. When I hear stories like my friend’s, I chuckle at the thought of showing up to job interviews with an empty plum-tomato tin on my head.

I’ve heard the sales argument before. As scientists, we need to ‘sell our work’ (or, my favourite, we need to ‘make it sexy’). I get what that means; we need to highlight the most salient findings and implications of our work without hiding its caveats and limitations. Instead, we should turn them into mere blemishes under the flattering light of our discoveries.

But how should we go about that? Or, more importantly, how can we avoid overdoing it? When I hear that I need to sell my work, my grant proposal and even myself, I can’t shake the fear that if I push too hard, I will start sounding like the proverbial used-car salesman.

I decided to approach the question as any sensible scientist would: I googled ‘key steps to a successful sale’. A few tips caught my eye, because scientists tend to overlook them. They can be used when applying for a faculty position or research grant, or when pitching a paper to an academic journal.

First, be realistic and make sure that your product fits the needs of your target audience. Sales associates understand that they will not be able to sell everything to everyone. I don’t only mean trim the sub-par science; a tough sale won’t always be overcome by polishing your product. Even your most elegant science may not fly with an audience ‘not in the market’ for it, so be sure to pick your target wisely.

Second, a sales meeting is a conversation. All the tips I found stressed that the salesperson must listen to potential buyers to understand

their needs. It might seem obvious, but it got me thinking about how often we scientists make the process mostly about ourselves (my CV, my publications, my recommendation letters, me, me, me...). Some experts even advise forcing the conversation if it isn’t part of the

interview or application process. For example, they might arrange a series of phone interviews to get to know their ‘buyers’ — the hiring committee, the grant makers, the journal editors — before the interview or submission. But remember that looking up faculty members online doesn’t count as a conversation.

Finally, explain clearly what will happen after the sale. Buyers need to know how they will put you, the product, to use. Think of yourself as a new printer. Are you ‘upgradable’? Your prospective employers might want to know how easily you can scale your lab up or down, or move it between floors or buildings if necessary. They may also appreciate knowing that you’re ready to lecture on several subjects at a moment’s notice. Do you come with a ‘service contract’? Try to demonstrate that you can count on a network of collaborators in case of problems. What are your ‘consumables’? If your research involves the use of unconventional materials, make sure that you show that you have thought carefully about how to secure them. Do you come with ‘pre-installed drivers’? Be honest about what you need to get started. It’s best to tell your department about the particle accelerator you’ll need in your basement before the fleet of moving trucks arrives.

These universal sales principles won’t apply to every case, but they could come in handy, especially in a tough job market or funding climate. Of course, should everything else fail, you can always break out the car-salesman routine. Look the search-committee members squarely in the eye, give them your widest grin and ask, “Say, what will it take for me to get this job today?” ■

**Mariano A. Loza-Coll** is a postdoc in genetics at the Salk Institute for Biological Studies in La Jolla, California.



P. GILLIGAN



# TED AGONISTES

*A job for life.*

BY RAHUL KANAKIA

Why won't the message get through that I am important, so important, and that whenever people come around the office and see me wedged into the little desk right by the door, they should know that I am the one who is really in charge?

Didn't the boss once say to me: "We couldn't run this place without all of you"? But he really did not mean you in the plural but you in the singular, meaning just me, because all the other 'me's they've created were just created from me.

You people walk in and think, "Oh, it's just another Ted. And I'll treat him like I treat the other Teds. I'll smile at him. He dies for a smile. He's maladaptive in a way that perfectly adapts him for high-stress environments." I read the reports. They wrote the reports after talking to me. I'm not another Ted, I am *the* Ted.

The other Teds do not believe me. I see them everywhere now. They're phasing out the Bills, replacing them with Teds. So there are clones of me wedged into tiny desks behind the doors of half the offices in this building: by 2051 it'll be half the offices in the world, or so Xerogenesis' statements say. I am a stockholder in Xerogenesis. I bought one stock in the company, just so I would get their quarterly statements and their newsletter to shareholders. I used to go to their big meetings down on Amelia Island in Florida. I'd catheterize myself and drive down, 15 hours from Friday night to Saturday morning and take a quick nap in my car, then go to the meetings, but all the executives started getting confused and telling me to handle their appointments and shooing me away to the upstairs rooms where their Teds were working. So I don't go anymore.

You don't want to hear this, but I am the real Ted. My semen cells were the ones double-dipped and recombined to make fetuses of me. I am the real one. I was the one who grew up in that third-floor apartment in Omaha. I was the one who looked behind the couches every morning to see if there were little chunky messes to clean up before the stink of my mom's regurgitated

dinner seeped into every inch of the house.

I wasn't born in a crèche, though I wish I had been. I think I would have liked it. At least in the crèche, the Teds know they're important, they know they'll get taken care of, as long as they do their job and find all the messes before the shift manager sounds the alarm. I never had that assurance.

Of course I'm a Ted, I had myself registered with the agency and everything. Looking the way I do, it was the only way to get a job. No one wants to hire an unlicensed Ted: he might have gone bad since leaving the crèche. But I'm somehow less than a Ted. Because I was not subjected to quality-control scrutiny, I'm not eligible for Ted-Care. And employers won't give it, not to me. I can't talk with the other Teds about life in the crèches, or the sweaty adolescent

orgies inside the tight little blind spots in the network of video cameras. And I'm older than them, much older, and

more ugly, and they look at me like I am a stranger.

I'm less than them, when I should be more. But no one is interested in my life. Even nice Dr Avery, he lost his professorship and stays at home all the time now, except for a few meetings a year with Xerogenesis people. He won't even answer my e-mails. I understand. He's ashamed of me. But why? I'm a human being. I'm like them, like all of everyone. No one exploited me. No one enslaved me.

I signed those papers, and I got my pension. I'll get it soon enough. The money's in the account, I can see it. I'm not stupid, though, you know. That's a pretty common misconception. I'm not stupid. I can read the reports. I know that they don't think I will live past 65. But what do they know? They are not gods. They can't predict the future.

I eat right. I exercise. I'm slim and trim and I have just one goal. I'm going to live to be a hundred. We all are. We all have our pensions, guaranteed from birth. That's what the report said. We're going to live to be a hundred, and I'll be the first. I'll show the hundreds of thousands, the millions, the tens of millions, show them all the way.

They'll stop making us, but then it'll be too late. We'll come up from the underground, and oh ... not do anything much. I will not remake the world with my doing, I will only remake it with my being. We'll live to be a hundred, and sit at desks behind the doors of tiny rooms. And we'll scratch away at keyboards. But we'll own the desks, and the rooms and the keyboards. We'll own everything.

You're smiling at me, and I know that you're thinking about all the little Teds sitting perfectly quiet for hours at a time, knowing that if they make a move they definitely won't get to eat, and knowing that even if they don't move, they still might not get to eat.

But I will get to eat. I'll eat my fill. I know it. ■

Rahul Kanakia is an international development consultant based in Oakland, California.



JACEY

# C<sub>4</sub> grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland

Jack A. Morgan<sup>1</sup>, Daniel R. LeCain<sup>1</sup>, Elise Pendall<sup>2</sup>, Dana M. Blumenthal<sup>1</sup>, Bruce A. Kimball<sup>3</sup>, Yolima Carrillo<sup>2</sup>, David G. Williams<sup>4</sup>, Jana Heisler-White<sup>4</sup>, Feike A. Dijkstra<sup>1,5</sup> & Mark West<sup>1</sup>

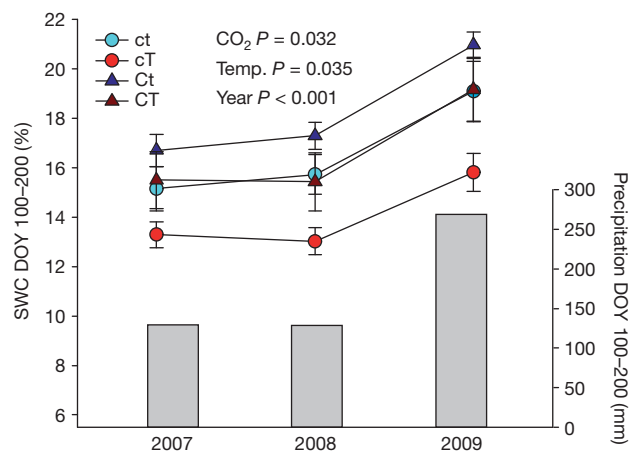
Global warming is predicted to induce desiccation in many world regions through increases in evaporative demand<sup>1–3</sup>. Rising CO<sub>2</sub> may counter that trend by improving plant water-use efficiency<sup>4,5</sup>. However, it is not clear how important this CO<sub>2</sub>-enhanced water use efficiency might be in offsetting warming-induced desiccation because higher CO<sub>2</sub> also leads to higher plant biomass, and therefore greater transpirational surface<sup>2,6,7</sup>. Furthermore, although warming is predicted to favour warm-season, C<sub>4</sub> grasses, rising CO<sub>2</sub> should favour C<sub>3</sub>, or cool-season plants<sup>8</sup>. Here we show in a semi-arid grassland that elevated CO<sub>2</sub> can completely reverse the desiccating effects of moderate warming. Although enrichment of air to 600 p.p.m.v. CO<sub>2</sub> increased soil water content (SWC), 1.5/3.0 °C day/night warming resulted in desiccation, such that combined CO<sub>2</sub> enrichment and warming had no effect on SWC relative to control plots. As predicted, elevated CO<sub>2</sub> favoured C<sub>3</sub> grasses and enhanced stand productivity, whereas warming favoured C<sub>4</sub> grasses. Combined warming and CO<sub>2</sub> enrichment stimulated above-ground growth of C<sub>4</sub> grasses in 2 of 3 years when soil moisture most limited plant productivity. The results indicate that in a warmer, CO<sub>2</sub>-enriched world, both SWC and productivity in semi-arid grasslands may be higher than previously expected.

Grass-dominated, dry rangelands account for over 30% of Earth's terrestrial surface<sup>9,10</sup> and provide most of the forage for the world's domestic livestock. Among the most important of these include the vast Central Asian steppes, North American prairies, Australian rangelands, plus extensive grazing lands of Africa, South America and the Mediterranean. Productivity of these lands, which is under increasing pressure with human population growth, is primarily limited by water<sup>9,11</sup>. Soil water availability is driven by complex interactions between precipitation, temperature<sup>1,11</sup>, ambient CO<sub>2</sub> (ref. 4) and soil properties, rendering rangelands and the livestock and cultures they support particularly susceptible to climate change and rising CO<sub>2</sub>. Higher CO<sub>2</sub> concentrations induce stomatal closure, which reduces leaf transpiration and increases plant water-use efficiency and SWC<sup>4</sup>. Despite agreement among global climate models that climate change will induce drought-like conditions in several rangeland-dominated world regions<sup>1,2</sup>, considerable uncertainty remains as to how stomatal closure from CO<sub>2</sub> might interact with increased temperature to influence stand-level evapotranspiration, soil water dynamics and plant productivity. Many believe that CO<sub>2</sub>-induced reductions in transpiration at the leaf level will be largely offset at the canopy level by increases in leaf area<sup>2,6,7,12</sup>.

Most rangelands are dominated by graminoids, and contain a mixture of the two major photosynthetic pathway classes of plants, C<sub>3</sub> and C<sub>4</sub>. Today, C<sub>4</sub> grasses are most common in grasslands and savannas within 40° of the Equator<sup>8</sup>, and are predicted to become more abundant with increasing temperature<sup>13</sup>. However, C<sub>3</sub> plants, owing to their CO<sub>2</sub>-limited photosynthetic metabolism, may become more competitive in future CO<sub>2</sub>-enriched rangelands<sup>8,14</sup>. The relative abundances

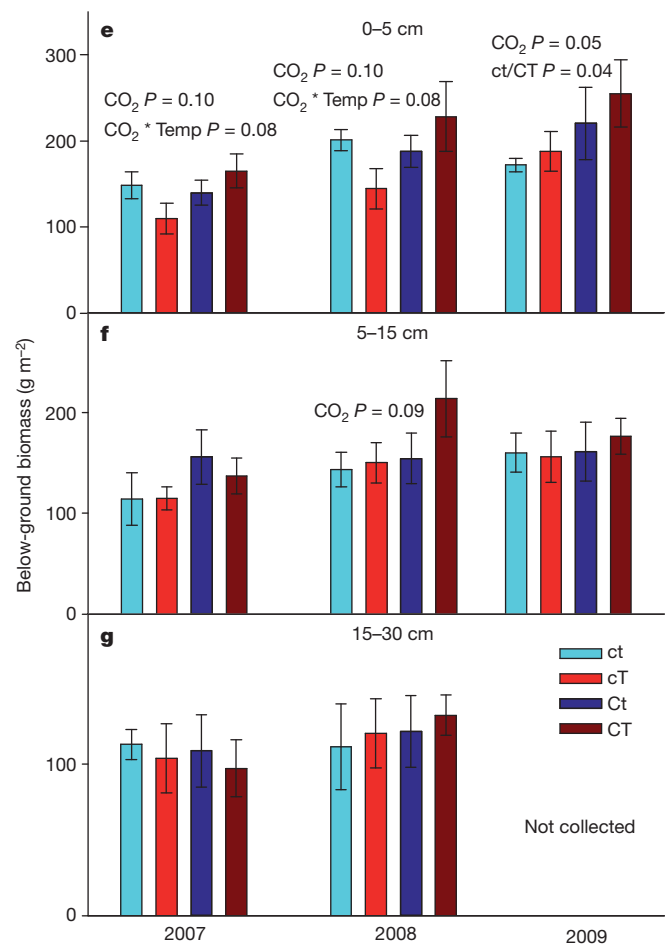
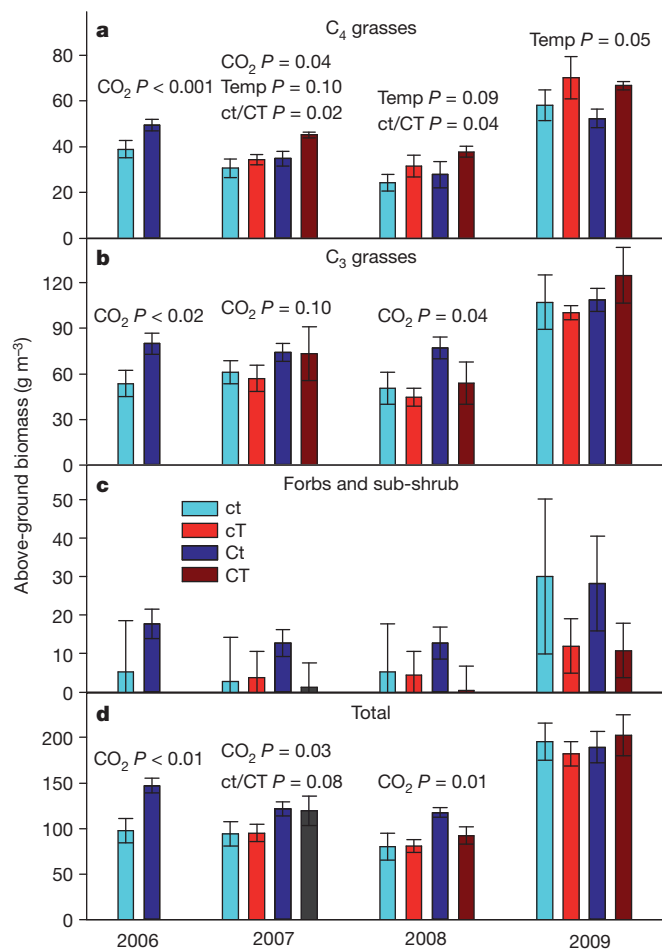
of these two photosynthetic types govern spatial and temporal aspects of water and nutrient cycling, net primary production and plant–animal interactions<sup>13,15,16</sup>, yet there is considerable uncertainty about which group will be favoured under future, warmer, CO<sub>2</sub>-enriched conditions.

In the Prairie Heating and CO<sub>2</sub> Enrichment (PHACE) experiment, we evaluated the responses of native mixed-grass prairie to 1 year of CO<sub>2</sub> enrichment (2006) (from present ambient (385 p.p.m.v.) to elevated (600 p.p.m.v.) CO<sub>2</sub>, treatments c and C respectively), followed by 3 more years (2007–2009) of combined CO<sub>2</sub> enrichment and warming (present ambient, and elevated, temperatures (1.5/3.0 °C warmer canopy day/night), treatments t and T respectively) using free-air CO<sub>2</sub> enrichment (FACE) technology for CO<sub>2</sub> (ref. 17) and T-FACE technology for warming<sup>18</sup> (Supplementary Appendix I). CO<sub>2</sub> enrichment and warming had opposing effects on soil water balance. Elevated CO<sub>2</sub> (Ct) increased annual SWC (17.3% averaged over 2007–2009), whereas warming reduced it (13.1%), such that there was no difference between control (ct; 15.5%) and CO<sub>2</sub>-enriched and warmed plots (CT; 15.6%) (Supplementary Appendix II). The same pattern of compensating treatment effects on SWC was observed in the early growing season (Fig. 1), the time of year when SWC most controls annual productivity in this grassland<sup>19</sup>. These results are similar to but more consistent than those reported in a semi-arid Australian temperate grassland<sup>20</sup>, and illustrate that the water conservation effects



**Figure 1 | Responses of SWC to CO<sub>2</sub> and warming.** Average and s.e.m. (error bars) of volumetric SWC (5–25 cm depth) for plots exposed to present-day ambient CO<sub>2</sub> and temperature (ct), 1.5/3 °C day/night warming (cT), 600 p.p.m.v. CO<sub>2</sub> (Ct), and 600 p.p.m.v. CO<sub>2</sub> and 1.5/3 °C day/night warming (CT) (five replications per treatment). Data are averaged over days of year (DOY) 100–200, the early- to mid-growing season when soil water most limits productivity. Precipitation amounts for this same period are also presented. Significance ( $P \leq 0.05$ ) for main effects and year are given in the figure.

<sup>1</sup>USDA-ARS, Rangeland Resources Research Unit and Northern Plains Area, Fort Collins, Colorado 80526, USA. <sup>2</sup>Department of Botany and Program in Ecology, University of Wyoming, Laramie, Wyoming 82071, USA. <sup>3</sup>US Arid-Land Agricultural Research Center, USDA, Agricultural Research Service, Maricopa, Arizona 85238, USA. <sup>4</sup>Departments of Botany, Renewable Resources, and Program in Ecology, University of Wyoming, Laramie, Wyoming 82071, USA. <sup>5</sup>Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Sydney, New South Wales 2006, Australia.



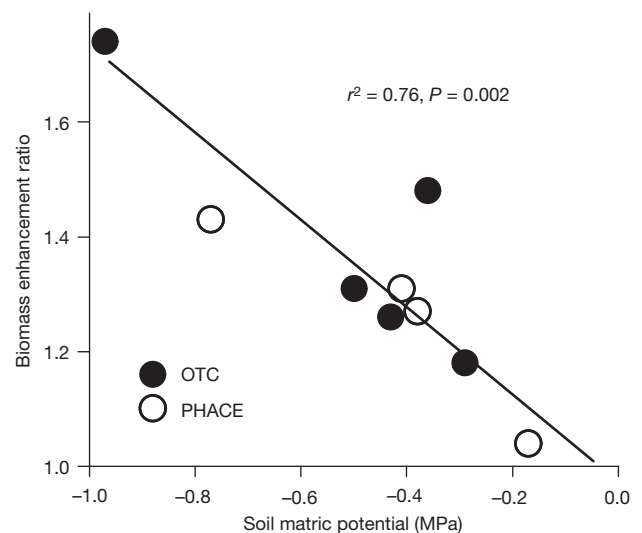
**Figure 2 | Plant biomass responses to CO<sub>2</sub> and warming.** Treatment effects (details in Fig. 1) on biomass in mid-July. AGB and s.e.m. (error bars) are given for C<sub>4</sub> (a) and C<sub>3</sub> grasses (b), forbs and sub-shrub (c) and total plant community (d). Below-ground biomass and s.e.m. (error bars) is given for soil depths of 0–5

(e), 5–15 (f) and 15–30 (g) cm. Results from 2006 include only the CO<sub>2</sub> treatment. Significant contrasts between present-day (ct) and future (CT) conditions are represented above histograms as ct/CT.  $P$  values are given for analyses of main effects, interactions and the ct/CT contrast when  $P \leq 0.10$ .

of elevated CO<sub>2</sub> can completely cancel the desiccating effects of moderately warmer temperatures.

Exposure of the prairie to 600 p.p.m.v. CO<sub>2</sub> increased peak total above-ground biomass (AGB, an estimate of above-ground net primary productivity) by an average 33% (Fig. 2d) in the first 3 years of the experiment when annual precipitation amounts were within 7% of the site 132-year average of 388 mm. We attribute this strong, positive response of AGB to CO<sub>2</sub> in this semi-arid grassland to the higher SWC in CO<sub>2</sub>-enriched plots<sup>4</sup>. CO<sub>2</sub> enrichment had no effect on AGB in 2009, a relatively wet year for the site when annual precipitation was 17% higher than the long-term mean (Supplementary Appendix II). We suspect that higher SWC in 2009 (Fig. 1) minimized the potential water-relations benefit of CO<sub>2</sub> enrichment on plant productivity.

To evaluate further the effect of CO<sub>2</sub> on SWC and plant productivity, we examined how responses of AGB to increasing CO<sub>2</sub> varied with early-season soil matric potential ( $\psi_m$ ) averaged during periods of active growth ( $\psi_m \geq -1.5$  MPa). The ratio of AGB of plants exposed to 600 p.p.m.v. CO<sub>2</sub> to that of plants exposed to present-day CO<sub>2</sub> (biomass enhancement ratio) was used for this evaluation. We incorporated results from another CO<sub>2</sub> enrichment experiment in the Colorado shortgrass steppe to include the other semi-arid grassland of the western Great Plains<sup>21</sup> (Supplementary Appendix I, Supplementary Fig. 1). A strong negative relationship was observed between  $\psi_m$  and biomass enhancement ratio (Fig. 3), which we attribute to (1) the enhancement of water use efficiency under elevated CO<sub>2</sub> and increasing water stress<sup>4,5</sup>, and (2) the use of a physiologically relevant environmental driver ( $\psi_m$  rather than SWC; Supplementary Appendix I, soil water conversions)



**Figure 3 | Response of biomass enhancement ratio to soil matric potential.** Effects of early-season (DOY 100–200) rooting zone soil matric potential ( $\psi_m$ ) on biomass enhancement ratio, the ratio of mid-July harvested above-ground plant biomass in CO<sub>2</sub>-enriched plots divided by plant biomass from ambient CO<sub>2</sub> plots. Data are from the PHACE experiment (4 years' data,  $n = 4$ ), and from a previous open top chamber CO<sub>2</sub> enrichment experiment (5 years' data,  $n = 5$ ) conducted on Colorado shortgrass steppe<sup>21</sup>. For further details, see Supplementary Appendix I, soil water conversions.



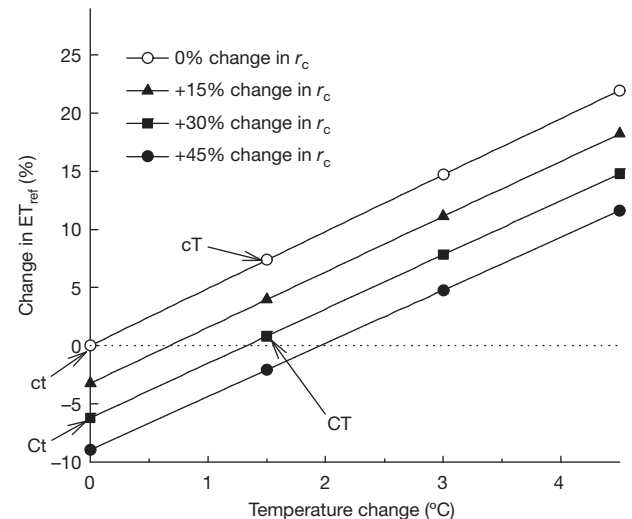
averaged over periods of physiological activity. This robust relationship from two different ecosystems suggests that CO<sub>2</sub> will increase plant productivity most when plants are water-limited yet still active.

The positive effect of CO<sub>2</sub> on production was also apparent in below-ground plant biomass. However, because root turnover in semi-arid perennial grasslands occurs every 5–7 years (ref. 22), we did not detect a significant ( $P \leq 0.05$ ) effect on roots until the fourth year of our experiment (2009), when below-ground plant biomass in the 0–5 cm depth layer was 32% higher in CO<sub>2</sub>-enriched plots (Fig. 2e).

Warming has the potential to increase production by extending the duration of the growing season, and by increasing nitrogen uptake, efficiency of nitrogen use<sup>23</sup> and biological activity. In our experiment, however, warming by itself did not significantly affect total AGB (Fig. 2d) or below-ground plant biomass (Fig. 2e–g), despite increased plant nitrogen content<sup>24</sup>. Warming-induced desiccation in this semi-arid mixed-grass prairie (Fig. 1) and the preponderance of cool-season C<sub>3</sub> grasses (55% of AGB) probably limited its ability to respond positively to warming.

Although water was a dominant driver in the responses of both C<sub>3</sub> and C<sub>4</sub> grasses to CO<sub>2</sub> and temperature, there was also evidence of unique functional group responses. C<sub>3</sub> grass AGB was 34% greater in CO<sub>2</sub>-enriched compared with ambient plots from 2006 to 2008 ( $P < 0.05$  in 2006 and 2008;  $P = 0.1$  in 2007) (Fig. 2b), but did not respond to warming. In contrast, AGB of C<sub>4</sub> grasses was 28% greater in CO<sub>2</sub>-enriched plots in the first 2 years, but was also greater under warming ( $P \leq 0.10$ ), most notably the wet year (2009) when C<sub>4</sub> AGB was 23% greater in heated plots ( $P = 0.05$ ) (Fig. 2a). Comparing present (ct) to future (CT) conditions, C<sub>3</sub> grasses did not respond, but C<sub>4</sub> AGB significantly increased with the combination of warming and CO<sub>2</sub> enrichment in 2007 and 2008 (Fig. 2). These results are consistent with our physiological understanding of C<sub>3</sub> and C<sub>4</sub> plant functional types<sup>8,13</sup>, but they demonstrate for the first time in a field experiment a potential advantage for C<sub>4</sub> grasses in a CO<sub>2</sub>-enriched and warmer future. Further research will be needed to determine the extent to which these results can be applied to other C<sub>3</sub> and C<sub>4</sub> grasses and grassland ecosystems (Supplementary Appendix III, C<sub>4</sub> versus C<sub>3</sub> responses). Small population sizes and high year-to-year variability limited our capacity to detect treatment effects on forbs and subshrubs (Fig. 2c).

Our results suggest that even with modest increases in plant growth (2007), rising atmospheric CO<sub>2</sub> may counter the pervasive desiccating effects of warming in semi-arid grassland ecosystems. The most likely explanation is a CO<sub>2</sub>-induced suppression of transpiration<sup>4</sup>. A modelling analysis using our site's weather data shows the effects of different hypothetical increases in canopy resistance to water loss ( $r_c$ ) such as might occur from stomatal closure, and different temperature increases on grass reference evapotranspiration rate (ET<sub>ref</sub>, Fig. 4), the evapotranspiration rate of non-stressed grassland. Differences in ET<sub>ref</sub> reflect not only the rate of evapotranspiration when soil water is readily available to plants, but also how long grass can continue to transpire and grow after major rain events before exhausting its soil water supply. At ambient CO<sub>2</sub>, a daytime warming of 1.5 °C (our cT treatment daytime temperature; roughly 90% of evapotranspiration occurs during daytime) is predicted to increase ET<sub>ref</sub> by 7%. At ambient temperature, an increase in  $r_c$  of 30% would be required to decrease ET<sub>ref</sub> by a similar amount, 6%. This condition is similar to our Ct treatment (Fig. 4); leaf gas exchange measurements on the dominant C<sub>3</sub> and C<sub>4</sub> grasses of this<sup>25</sup> and other<sup>26</sup> natural ecosystems suggest a 30–40% increase in stomatal resistance occurs in native grasses under CO<sub>2</sub>-enriched and non-stressed conditions. Thus, when daytime warming of 1.5 °C is combined with an increase of 30% in  $r_c$  (our CT treatment), the temperature effect and the CO<sub>2</sub> effect through increased stomatal resistance almost exactly offset one another, so that an increase of about only 1% in ET<sub>ref</sub> is predicted. This modelling result mirrors the observed offsetting effects of CO<sub>2</sub> and temperature on SWC (Figs 1 and Supplementary Figure 4). The analysis does not account for treatment



**Figure 4 | Percentage changes in ET<sub>ref</sub> for a grass surface as affected by temperature and changes in  $r_c$ .** Percentage changes in ET<sub>ref</sub> calculated using the American Society of Civil Engineers standardized evapotranspiration equation versus a range of temperature increases following Kimball<sup>28</sup>. Calculations used observed weather data from the PHACE project for the 1 April – 16 October 2007, 2008 and 2009 growing seasons. The total calculated ET<sub>ref</sub> for the three seasons was 2,490 mm (average of 5.1 mm d<sup>-1</sup>). Calculations were done for zero, +15, +30 and +45% changes in canopy resistance to water vapour ( $r_c$ ), as might be expected from CO<sub>2</sub>-induced stomatal closure. For further details, see Supplementary Appendix III, Supplementary Fig. 5.

effects on leaf area, although such differences were probably small compared with 30–40% increases in stomatal resistance we would expect under CO<sub>2</sub> enrichment (Supplementary Appendix III, evapotranspiration modelling). A protracted drought that depletes soil water stores will of course eventually eliminate these CO<sub>2</sub>-induced water savings; we have yet to experience such conditions in this experiment. Thus, although rising CO<sub>2</sub> may ameliorate the effects of warming-induced drought through plant stomatal closure, it is unlikely to eliminate severe and protracted droughts predicted for rangelands in regions like southwestern North America or the Mediterranean, where both substantive temperature increase and reduced precipitation are predicted<sup>1,2</sup>.

The particular responses reported here are to a single set of simulated global change conditions that are within the prediction possibilities of the Intergovernmental Panel on Climate Change for the latter half of this century<sup>27</sup>. The manipulations are not without their artefactual issues, which must be considered in scaling such information temporally and beyond the experimental site (Supplementary Appendix III, warming & CO<sub>2</sub> methodologies). Nevertheless, our results clearly illustrate the importance of compensating CO<sub>2</sub> and warming effects in semi-arid ecosystems. These contrasting, water-mediated effects must be accounted for in accurately predicting the susceptibility of such systems to climate change.

## METHODS SUMMARY

**Experiment.** The experiment was conducted in northern mixed-grass prairie west of Cheyenne, Wyoming, USA. It had a factorial combination of two levels of CO<sub>2</sub> (385 p.p.m.v. and elevated 600 p.p.m.v.), and two temperature (present ambient, and elevated (1.5/3.0 °C warmer day/night)) regimes, with five replications each. Treatments were randomly assigned to 20 3.3-m diameter circular plots in two soil types (blocks). Details of the experimental site and set up are available elsewhere<sup>24</sup>.

AGB was measured by species in 2005 (the year before CO<sub>2</sub> treatments), in 2006 (the first year of CO<sub>2</sub> treatments) and from 2007 to 2009 (the years of combined elevated CO<sub>2</sub> and warming treatments) during mid-July. A metal wire grid divided into 24–25 cm × 25 cm quadrats (1.5 m<sup>2</sup> total) was placed over each plot and vegetation in every other quadrat (12 in total) was clipped to the crown, dried at 60 °C and weighed. Plant species were subsequently grouped into three functional groups for analyses: C<sub>4</sub> grasses, C<sub>3</sub> grasses, and Forbs and sub-shrub.

Beginning in 2007, three soil cores (30 cm deep, 5 cm diameter) were collected at about the same time as the AGB samples from each plot, divided into 0–5, 5–15 and 15–30 cm depths, and the three core samples composited into one sample per depth. Fine roots (less than 1 mm) were handpicked from the composited samples, and root fragments were washed, weighed and ash-corrected.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 17 December 2010; accepted 7 June 2011.

Published online 3 August 2011.

- Wang, G. Agricultural drought in a future climate: results from 15 global change models participating in the IPCC 4<sup>th</sup> assessment. *Clim. Dyn.* **25**, 739–753 (2005).
- Seager, R. & Vecchi, G. A. Greenhouse warming and the 21<sup>st</sup> century hydroclimate of southwestern North America. *Proc. Natl Acad. Sci. USA* **107**, 21277–21282 (2010).
- Woodhouse, C. A., Meko, D. M., MacDonald, G. M., Stahle, D. W. & Cook, E. R. A 1,200-year perspective of 21<sup>st</sup> century drought in southwestern North America. *Proc. Natl Acad. Sci. USA* **107**, 21283–21288 (2010).
- Morgan, J. A. *et al.* Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO<sub>2</sub>. *Oecologia* **140**, 11–25 (2004).
- Leakey, A. D. B. Rising atmospheric carbon dioxide concentration and the future of C<sub>4</sub> crops for food and fuel. *Proc. R. Soc. Lond. B* **276**, 2333–2343 (2009).
- Frelich, L. E. & Reich, P. B. Will environmental changes reinforce the impact of global warming on the prairie–forest border of central North America? *Front. Ecol. Environ.* **8**, 371–378 (2010).
- Piao, S. *et al.* Changes in climate and land use have a larger direct impact than rising CO<sub>2</sub> on global river runoff trends. *Proc. Natl Acad. Sci. USA* **104**, 15242–15247 (2007).
- Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C–4 photosynthesis, atmospheric CO<sub>2</sub> and climate. *Oecologia* **112**, 285–299 (1997).
- Asner, G. P., Elmore, A. J., Olander, L. P., Martin, R. E. & Harris, A. T. Grazing systems, ecosystem responses, and global change. *Annu. Rev. Environ. Resour.* **29**, 261–299 (2004).
- Suttie, J. M., Reynolds, S. G. & Batello, C. *Grasslands of the World* (FAO, 2005).
- Noy-Meir, I. Desert ecosystems: environment and producers. *Annu. Rev. Ecol. Syst.* **4**, 25–51 (1973).
- McNaughton, K. G. & Jarvis, P. G. Effects of spatial scale on stomatal control of transpiration. *Agric. For. Meteorol.* **54**, 279–301 (1991).
- Epstein, H. E. *et al.* The relative abundance of three plant functional types in temperate grasslands and shrublands of North and South America: effects of projected climate change. *J. Biogeogr.* **29**, 875–888 (2002).
- Polley, H. W. Implications of rising atmospheric carbon dioxide concentration for rangelands. *J. Range Mgmt* **50**, 562–577 (1997).
- Semmartin, M., Aguiar, M. R., Distel, R. A., Moretto, A. S. & Ghersa, C. M. Litter quality and nutrient cycling affected by grazing-induced species replacements along a precipitation gradient. *Oikos* **107**, 148–160 (2004).
- Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K. & DeJong, D. D. NDVI, C<sub>3</sub> and C<sub>4</sub> production, and distributions in Great Plains grassland land cover classes. *Ecol. Appl.* **7**, 59–78 (1997).
- Miglietta, F. *et al.* Free-air CO<sub>2</sub> enrichment (FACE) of a poplar plantation: the POPFACE fumigation system. *New Phytol.* **150**, 465–476 (2001).
- Kimball, B. A. *et al.* Infrared heater arrays for warming ecosystem field plots. *Glob. Change Biol.* **14**, 309–320 (2008).
- Derner, J. D. & Hart, R. H. Grazing-induced modifications to peak standing crop in northern mixed-grass prairie. *Rangeland Ecol. Mgmt* **60**, 270–276 (2007).
- Hovenden, M. J. *et al.* Influence of warming on soil water potential controls seedling mortality in perennial but not annual species in a temperate grassland. *New Phytol.* **180**, 143–152 (2008).
- Morgan, J. A., Milchunas, D. G., LeCain, D. R., West, M. & Mosier, A. R. Carbon dioxide enrichment alters plant community structure and accelerates shrub growth in the shortgrass steppe. *Proc. Natl Acad. Sci. USA* **104**, 14724–14729 (2007).
- Milchunas, D. G., Morgan, J. A., Mosier, A. R. & LeCain, D. R. Root dynamics and demography in shortgrass steppe under elevated CO<sub>2</sub>, and comments on minirhizotron methodology. *Glob. Change Biol.* **11**, 1837–1855 (2005).
- Luo, Y., Sherry, R., Zhou, X. & Wan, S. Terrestrial carbon-cycle feedback to climate warming: experimental evidence on plant regulation and impacts of biofuel feedstock harvest. *Glob. Change Biol. Bioenergy* **1**, 62–74 (2009).
- Dijkstra, F. A. *et al.* Contrasting effects of elevated CO<sub>2</sub> and warming on nitrogen cycling in a semiarid grassland. *New Phytol.* **187**, 426–437 (2010).
- LeCain, D. R., Morgan, J. A., Mosier, A. R. & Nelson, J. A. Soil and plant water relations determine photosynthetic responses of C<sub>3</sub> and C<sub>4</sub> grasses in a semi-arid ecosystem under elevated CO<sub>2</sub>. *Ann. Bot. (Lond.)* **92**, 41–52 (2003).
- Wand, S. J. E., Midgley, G. F., Jones, M. H. & Curtis, P. S. Responses of wild C<sub>4</sub> and C<sub>3</sub> grass (Poaceae) species to elevated atmospheric CO<sub>2</sub> concentrations: a meta-analytic test of current theories and perceptions. *Glob. Change Biol.* **5**, 723–741 (1999).
- Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC* (eds Solomon, S. *et al.*) 747–845 (Cambridge Univ. Press, 2007).
- Kimball, B. A. in *Irrigation of Agricultural Crops (Agronomy Monograph No. 30)* 2nd edn (eds Lascano, R. J. & Sojka, R. E.) 627–653 (American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, 2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Smith for installation and operation of the PHACE experiment, E. Hardy for assistance in installation, A. Eden and C. Brooks for data collection and analysis, F. Miglietta for advice and help on installation of the FACE system, and R. Seager, A. Leakey, B. Cook and G. Wang for comments on the manuscript. The work was supported by the US Department of Agriculture–Agricultural Research Service Climate Change, Soils & Emissions Program, the US Department of Agriculture–Cooperative State Research, Education, and Extension Service Soil Processes Program (grant no. 2008-35107-18655), the US Department of Energy's Office of Science (Biological and Environmental Research) through the Western Regional Center of the National Institute for Climatic Change Research at Northern Arizona University, and the National Science Foundation (DEB no. 1021559). Mention of commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

**Author Contributions** J.A.M., E.P., D.M.B., B.A.K., D.G.W. and M.W. conceived the study. J.A.M., D.R.L., E.P., D.M.B., Y.C., D.G.W., J.H.-W. and F.A.D. performed the experiment. B.A.K. designed the warming system and conducted the evapotranspiration analysis. J.A.M. wrote the paper and the remaining authors contributed to the writing. Statistical analyses using SAS/STAT software, version 9.2, Proc GLIMMIX were performed by M.W. and J.A.M. The regression analyses using JMP software were performed by D.M.B. and J.A.M. Figures were developed by D.R.L.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.A.M. ([jack.morgan@ars.usda.gov](mailto:jack.morgan@ars.usda.gov)).

## METHODS

**Site description.** The experiment was conducted at the US Department of Agriculture-Agricultural Research Service High Plains Grasslands Research Station, west of Cheyenne, Wyoming, USA (41° 11' N, 104° 54' W), elevation 1930 m. The ecosystem is a northern mixed-grass prairie, with a plant community of about 55% cool-season C<sub>3</sub> grasses (mostly *Pascopyrum smthii* and *Hesperostipa comata*, both perennials), 25% warm-season C<sub>4</sub> grasses (almost exclusively the perennial *Bouteloua gracilis*), and 20% sedges, forbs and small shrubs. Annual precipitation is 384 mm, mean air temperatures are 17.5 °C in summer and −2.5 °C in winter. The average annual wind speed is 6 m s<sup>−1</sup>, with a growing season wind speed of 4.1 m s<sup>−1</sup>.

A 2.4-ha site, which had been grazed by cattle at least since 1974, was fenced to prevent cattle entering in 2005. Soils are an Ascalon variant loam (fine-loamy, mixed-mesic) at the north end of the study site, and an Altvan loam (fine-loamy over sandy, mixed-mesic) at the south end.

**Experimental layout.** The experiment has a factorial combination of two levels of CO<sub>2</sub> (present ambient (385 p.p.m.v.), and elevated (600 p.p.m.v.), c and C respectively), and two temperature (present ambient, and elevated (1.5/3.0 °C warmer day/night), t and T respectively) regimes, with five replications making 20 experimental plots. Assuming an annual increase of CO<sub>2</sub> concentration of 2.3 p.p.m.v., our treatment combinations represent CO<sub>2</sub> and temperatures expected near the end of this century. We chose differential day/night temperatures based on the Fourth Intergovernmental Panel on Climate Change assessment, which suggests that daily minimal temperatures will increase faster than maximal temperatures<sup>27</sup>. Furthermore, the equilibrium global mean surface air temperature for a doubling of atmospheric CO<sub>2</sub> is likely to lie between 2 and 4.5 °C. For our target CO<sub>2</sub> concentration of 600 p.p.m.v., our average day/night temperature of 2.25 °C lies at the upper third of a range we calculate to be between approximately 1.1–2.5 °C. The four treatment combinations were designated as follows: ct, ambient CO<sub>2</sub>, non-heated; Ct, elevated CO<sub>2</sub>, non-heated; cT, ambient CO<sub>2</sub>, heated; CT, elevated CO<sub>2</sub>, heated. Treatments were randomly assigned to 20 plots in northern and southern soil-type field blocks. Details are available elsewhere<sup>24</sup>, including the FACE<sup>17</sup> and infrared warming<sup>18</sup> apparatus used to alter ambient CO<sub>2</sub> and temperature (see also Supplementary Appendix I).

**Experimental rings and microclimate.** To accommodate the FACE system, experimental plots were circular with a diameter of 3.3 m (area of 8.6 m<sup>2</sup>). Before instrumentation was installed, hexagonal trenches 60 cm deep were dug around the circumference of each plot and a plastic barrier installed to isolate treated plots hydraulically from outside non-treated soils. This hydraulic separation helped maintain any water-relations effects of the treatments.

Most of the plot area was maintained as native northern mixed prairie and was partitioned into soil and plant sampling areas, plus sections for other measurements, including SWC and air and soil temperature. Volumetric SWC was

measured daily using frequency domain reflectometry sensors at 10 and 20 cm soil depths (Sentek EnviroSMART sensors, Sentek Sensor Technologies) placed approximately 75 cm from the ring centres. Daily total precipitation was recorded with a rain gauge (Onset corp. S-RGA-M002).

**Plant biomass.** AGB was measured by species in 2005 (the year before CO<sub>2</sub> treatments), in 2006 (the first year of CO<sub>2</sub> treatments) and from 2007–2009 (the years of combined elevated-CO<sub>2</sub> and warming treatments) during mid-July. Mid-July is the approximate time of peak seasonal AGB, and provides a good estimate of above-ground net primary production. A metal wire grid divided into 24–25 cm × 25 cm quadrats (1.5 m<sup>2</sup> total) was placed over each plot (all sampled vegetation being within a metre of the ring centre), and vegetation in every other quadrat (12 in total) was clipped to the crown, dried at 60 °C and weighed. This defoliation protocol removed 50% of the green vegetation and represented the grazing removal for these grasslands. To correct for initial plot differences in species abundance, the 2005 pre-treatment harvest data were used to calculate overall means of each plant species' biomass across all 20 plots, plus the deviation for each species from that mean in every plot. These deviations were either added to or subtracted from individual species' biomass data collected in all plots for the subsequent treatment years, 2006–2009, thereby correcting for initial plot differences in plant species abundance. Plant species were subsequently grouped into three functional groups for analyses: C<sub>4</sub> grasses, C<sub>3</sub> grasses and sedges, and forbs and sub-shrub.

Beginning in 2007, near the time of the peak standing AGB harvest, three soil cores (30 cm deep, 5 cm diameter) were collected from each plot, divided into 0–5, 5–15 and 15–30 cm depths, and composited into one sample per depth. Fine roots (less than 1 mm) were handpicked from the composited samples, and root fragments were washed, weighed and ash corrected.

**Statistical analyses.** Data analysis used SAS/STAT software, version 9.2, Proc GLIMMIX, © 2002–2008 SAS Institute. Mean AGB and below-ground plant biomass were compared among all four treatments (five replications each) and years by fitting a repeated-measures general linear model to biomass. Plots within treatments represented subjects on which repeated measures were taken. Treatments and year were fitted as fixed-effects. Contrasts were constructed to estimate and test the effects of CO<sub>2</sub>, warming and their interaction on biomass within years, owing to climatic differences among the four years. We hypothesized that increasing CO<sub>2</sub> would increase plant biomass (one-tailed test), but that warming would have no net effect on biomass (two-tailed test). Analyses of mean SWC (averaged annually for days of year 100–200) were conducted across years 2007–2009 when both CO<sub>2</sub> and warming treatments were operational. Regression curves of SWC and  $\psi_m$  with biomass enhancement ratio using data from the PHACE and open top chamber experiments were computed using JMP software, version 8.0.1, 2009 SAS Institute. Throughout, a *P* value of ≤ 0.05 is the standard for indicating significance, although marginal significance (*P* ≤ 0.10) is reported in figures, and sometimes commented on in the text as deemed appropriate.

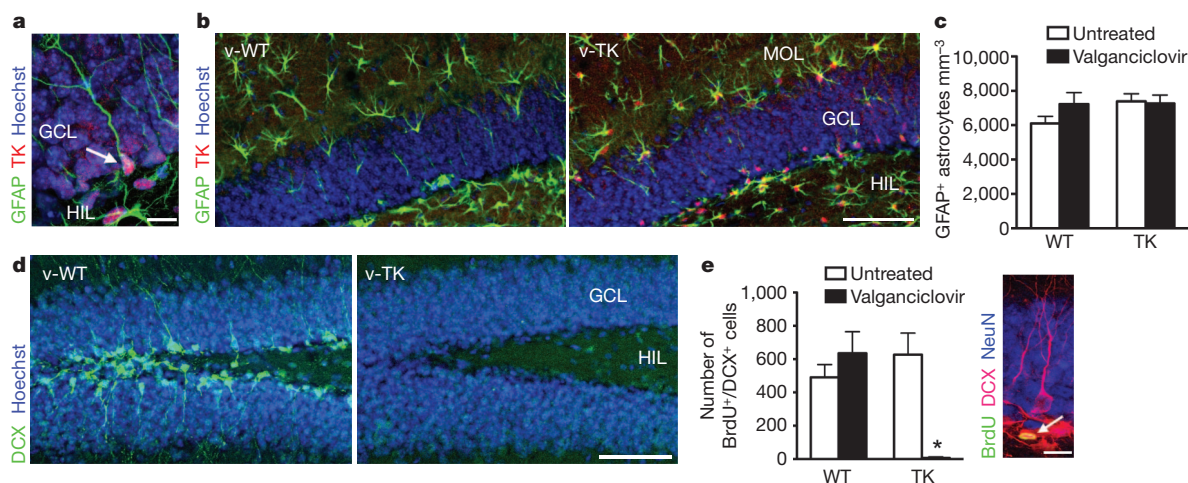


# Adult hippocampal neurogenesis buffers stress responses and depressive behaviour

Jason S. Snyder<sup>1</sup>, Amélie Soumier<sup>1</sup>, Michelle Brewer<sup>1</sup>, James Pickel<sup>1</sup> & Heather A. Cameron<sup>1</sup>

Glucocorticoids are released in response to stressful experiences and serve many beneficial homeostatic functions. However, dysregulation of glucocorticoids is associated with cognitive impairments and depressive illness<sup>1,2</sup>. In the hippocampus, a brain region densely populated with receptors for stress hormones, stress and glucocorticoids strongly inhibit adult neurogenesis<sup>3</sup>. Decreased neurogenesis has been implicated in the pathogenesis of anxiety and depression, but direct evidence for this role is lacking<sup>4,5</sup>. Here we show that adult-born hippocampal neurons are required for normal expression of the endocrine and behavioural components of the stress response. Using either transgenic or radiation methods to inhibit adult neurogenesis specifically, we find that glucocorticoid levels are slower to recover after moderate stress and are less suppressed by dexamethasone in neurogenesis-deficient mice than intact mice, consistent with a role for the hippocampus in regulation of the hypothalamic–pituitary–adrenal (HPA) axis<sup>6,7</sup>. Relative to controls, neurogenesis-deficient mice also showed increased food avoidance in a novel environment after acute stress, increased behavioural despair in the forced swim test, and decreased sucrose preference, a measure of anhedonia. These findings identify a small subset of neurons within the dentate gyrus that are critical for hippocampal negative control of the HPA axis and support a direct role for adult neurogenesis in depressive illness.

Functional granule neurons are generated in the hippocampus throughout life by a multistep process that begins with glial fibrillary acidic protein (GFAP)-expressing radial cell precursors<sup>8,9</sup>. To investigate the role of adult neurogenesis in hippocampal function, we created mice that express herpes simplex virus thymidine kinase (TK) under control of the GFAP promoter (TK mice; Fig. 1a). Thymidine kinase renders mitotic cells sensitive to the antiviral drug valganciclovir but spares post-mitotic cells<sup>10</sup>. Stellate astrocytes and radial neuronal precursor cells both express GFAP, and both cell types express thymidine kinase in the transgenic mice. However, the number of astrocytes is unaltered by treatment with valganciclovir in TK mice (Fig. 1b, c), consistent with a lack of cell proliferation in astrocytes in the adult brain<sup>11</sup>. In contrast, immature neurons expressing doublecortin (DCX) were virtually eliminated in the dentate gyrus of valganciclovir-treated TK (v-TK) mice (Fig. 1d and Supplementary Fig. 1). One-day-old neuronal progenitors, identified with DCX and the cell-cycle marker bromodeoxyuridine (BrdU), were reduced in v-TK mice by 99% relative to control conditions (Fig. 1e). Transgene expression alone and valganciclovir treatment of wild-type (WT) mice had no effect (Fig. 1e). v-TK mice showed weight gain comparable to v-WT mice (Supplementary Fig. 2), and histopathological examination found no abnormalities in the small intestine or submucosal or myenteric plexuses (not shown), indicating that the gastrointestinal effects described in another strain expressing thymidine kinase under the



**Figure 1 | GFAP-TK mice show specific inhibition of adult neurogenesis.** **a**, Confocal image of endogenous GFAP and transgenic thymidine kinase expression in a radial precursor cell in the dentate gyrus (arrow). **b**, Confocal photographs of valganciclovir-treated mice show GFAP<sup>+</sup> astrocytes in the hilus and molecular layer in both genotypes, despite strong thymidine kinase expression in all GFAP-expressing cells in TK mice. **c**, Valganciclovir treatment did not affect numbers of GFAP<sup>+</sup> astrocytes (genotype effect  $F_{1,20} = 1.7$ ,  $P = 0.2$ ; valganciclovir effect  $F_{1,20} = 1.0$ ,  $P = 0.3$ ; interaction  $F_{1,20} = 1.5$ ,  $P = 0.2$ ;  $n = 6$  per group), confirming the expectation that valganciclovir does not kill astrocytes, which are post-mitotic in the adult. **d**, Confocal photographs

of dentate gyrus doublecortin (DCX) immunostaining in mice treated with valganciclovir (v-WT and v-TK) for 4 weeks. DCX<sup>+</sup> young neurons are abundant in v-WT mice but absent in v-TK mice. **e**, The number of BrdU<sup>+</sup>/DCX<sup>+</sup> young neurons was unaltered in v-WT mice but reduced by 99% in v-TK mice (genotype effect  $F_{1,20} = 20$ ,  $P = 0.0002$ ; valganciclovir effect  $F_{1,20} = 19$ ,  $P = 0.0003$ ; interaction  $F_{1,20} = 40$ ,  $P < 0.0001$ ; \* $P < 0.001$  versus untreated TK and v-WT;  $n = 6$  per group). Inset shows example of 1-day-old BrdU<sup>+</sup>/DCX<sup>+</sup> neuron (arrow). Error bars, s.e.m.; scale bars, 10  $\mu$ m in **a**, **e**, 100  $\mu$ m in **b**, **d**. MOL, molecular layer; GCL, granule cell layer; HIL, hilus.

<sup>1</sup>National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA.

GFAP promoter<sup>8</sup> are absent in this line of mice. Taken together, these data indicate that v-TK mice have a specific loss of adult neurogenesis without detectable effects on astrocytes or general health.

The hippocampus provides negative control of the HPA axis<sup>6,7</sup>, but the circuitry involved is not well understood. Because adult hippocampal neurogenesis is highly sensitive to stress and glucocorticoids<sup>3</sup>, we hypothesized that adult neurogenesis may be important for hippocampal regulation of the HPA axis. We therefore examined serum levels of corticosterone, the predominant rodent glucocorticoid, in several conditions that activate the HPA axis. Neurogenesis-deficient v-TK and control v-WT mice had equivalent levels of corticosterone at the onset of both the light and the dark phase (Supplementary Fig. 3), indicating that adult-born neurons are not required for normal circadian fluctuation of glucocorticoids. v-WT and v-TK mice also had similar corticosterone levels after exposure to a novel environment (Supplementary Fig. 3), a mild stressor, consistent with previous findings<sup>12</sup>.

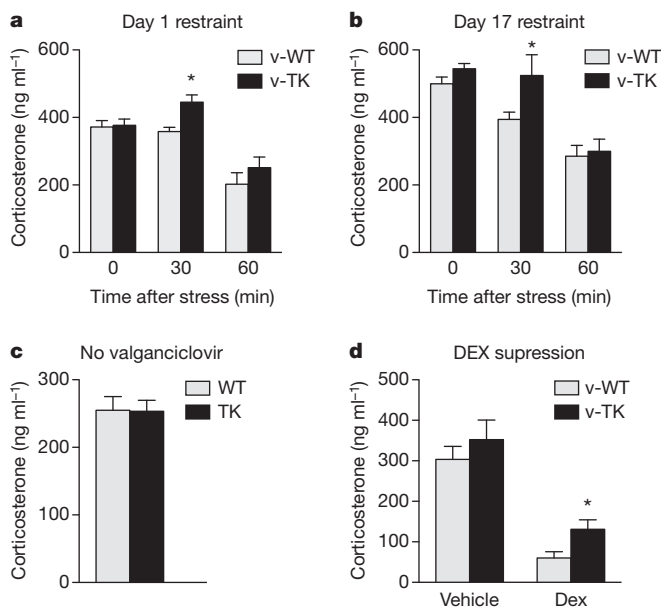
To test the response to, and recovery from, a moderate psychological stressor, we subjected mice to 30 min of restraint and measured corticosterone 0, 30 or 60 min later (Fig. 2a). v-TK mice and v-WT mice had similar levels of corticosterone immediately after termination of the stressor. However, v-TK mice had elevated corticosterone relative to v-WT mice 30 min after stress, suggesting impaired negative feedback control of glucocorticoid release similar to that observed in mice with complete loss of glucocorticoid receptors in the forebrain<sup>13</sup>. To investigate whether the hypersecretion of glucocorticoids habituates, mice were subjected to daily restraint for 16 additional days, and corticosterone was measured on the final day. As on the first day of restraint, v-TK mice had elevated corticosterone relative to v-WT mice (Fig. 2b). Control TK mice, which were never treated with valganciclovir and therefore had normal levels of neurogenesis, had levels of corticosterone identical to untreated WT mice 30 min after restraint (Fig. 2c). Therefore, corticosterone

hypersecretion in neurogenesis-deficient mice is not caused by non-specific or insertion site effects of the thymidine kinase transgene.

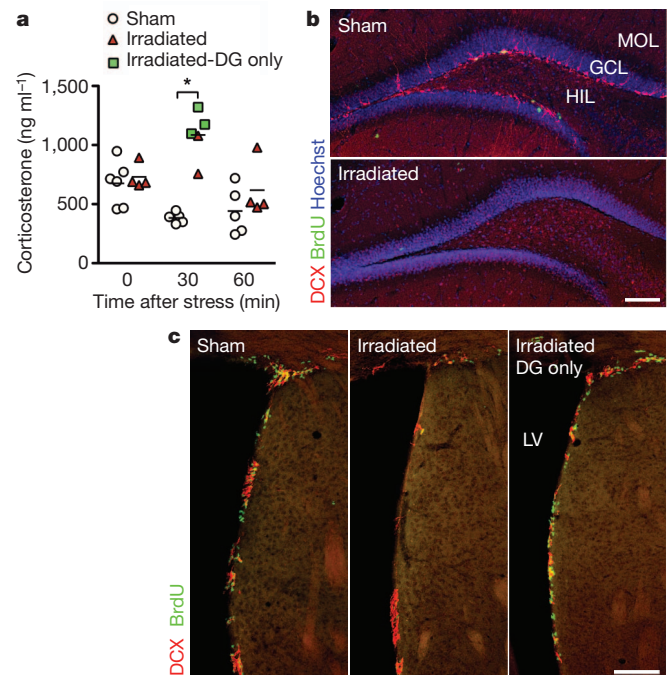
The experiments above indicate that adult neurogenesis regulates the endocrine stress response, but they do not pinpoint the brain region involved. Hippocampal damage primarily alters the response to psychological stressors such as restraint, which produce fear without causing a direct threat to well-being, but does not typically affect responses to physical stressors, such as hypoxia, haemorrhage, inflammation or anaesthesia<sup>7</sup>. Consistent with HPA regulation at the level of the hippocampus, v-WT and v-TK mice showed comparable elevations in corticosterone after exposure to isoflurane anaesthesia (Supplementary Fig. 4). Additionally, in agreement with the lack of global HPA dysregulation, we found no evidence for reduced cell birth in the hypothalamic paraventricular nucleus of v-TK mice (Supplementary Fig. 5).

To test the contribution of adult neurogenesis in other regions directly, we exploited the spatial specificity of X-irradiation to reduce hippocampal neurogenesis while sparing neurogenesis in the subventricular zone<sup>12,14</sup>, a source of GFAP<sup>+</sup> neuronal precursors added to the olfactory bulb throughout adulthood<sup>8,15</sup>. Irradiated mice had elevated corticosterone during recovery from restraint stress (Fig. 3), replicating the initial finding with an independent ablation method. In addition, there was no relationship between the extent of subventricular zone neurogenesis inhibition and the corticosterone response in irradiated mice (Fig. 3). Therefore, the most parsimonious interpretation of our results is that inhibition of adult neurogenesis in the dentate gyrus leads to hypersecretion of glucocorticoids in response to stress.

A functional link between new neurons and anxiety/depression has been suggested by the demonstration that some antidepressant effects on behaviour are blocked by irradiation<sup>12,16–18</sup>. However, normal anxiety- and depressive-like behaviour in animals lacking neurogenesis has led



**Figure 2 | The glucocorticoid response to stress is increased in neurogenesis-deficient mice.** **a**, Restraint, a moderate psychogenic stressor, resulted in higher corticosterone in neurogenesis-deficient v-TK mice than in v-WT mice 30 min after the end of stress. **b**, The effect of restraint was still observed after repeated exposure to stress (for both days: genotype effect  $F_{1,65} > 5$ ,  $P < 0.05$ ; time effect  $F_{2,65} > 24$ ,  $P < 0.001$ ;  $*P < 0.05$  post hoc versus v-WT;  $n = 6$ –17 per group per time point). **c**, In untreated control mice, corticosterone levels 30 min after restraint stress were not different between WT and TK mice, indicating that altered glucocorticoid response to stress is not a non-specific effect of transgene expression ( $t_{30} = 0.1$ ,  $P = 0.95$ ;  $n = 16$  per group). **d**, Valganciclovir-treated v-TK mice show impaired dexamethasone suppression of corticosterone in response to restraint ( $*t_{13} = 2.5$ ,  $P = 0.03$ ;  $n = 7$ –8 per group). Error bars, s.e.m.



**Figure 3 | Increased stress response is not due to reduced neurogenesis in the subventricular zone.** **a**, Increased corticosterone response 30 min after restraint was confirmed in mice in which neurogenesis was reduced by irradiation (irradiation effect  $F_{1,50} = 2.0$ ,  $P = 0.16$ ; time effect  $F_{2,50} = 5.1$ ,  $P = 0.01$ ; irradiation  $\times$  time interaction  $F_{2,50} = 5.7$ ,  $P = 0.006$ ;  $*P < 0.01$  post hoc;  $n = 4$ –6 per group per time point). Green squares at the 30 min time point indicate corticosterone levels in irradiated mice that showed unaffected neurogenesis in the subventricular zone. **b**, Confocal images of BrdU<sup>+</sup> and DCX<sup>+</sup> cells in the dentate gyrus; neurogenesis was reduced in all irradiated mice. **c**, Confocal images of BrdU<sup>+</sup> and DCX<sup>+</sup> cells illustrating sparing of neurogenesis in the subventricular zone. Scale bars, 100  $\mu$ m. LV, lateral ventricle.

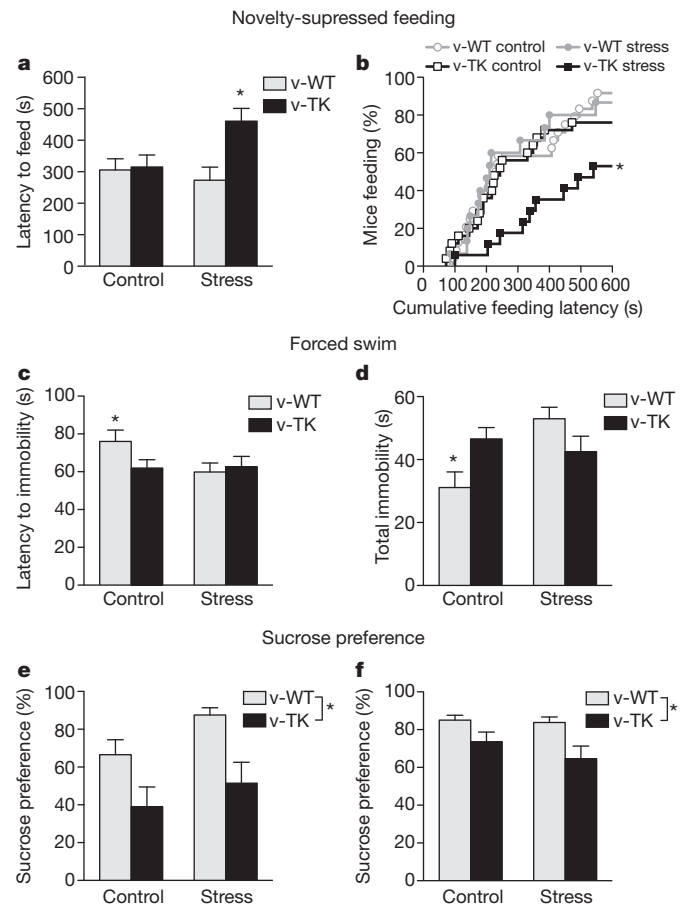
to speculation that impaired neurogenesis does not directly contribute to the aetiology of depression in adulthood<sup>4,5</sup>. Our findings above suggest that stress may be a key unexplored factor linking adult-born neurons to anxiety- and depressive-like behaviours.

To probe a potential interaction between stress, neurogenesis and depression, we first used the dexamethasone suppression test, which is commonly used to test HPA axis feedback in depressed patients. A subgroup of depressed patients, and mice that display depressive behaviours, show impaired inhibition of endogenous glucocorticoids by the synthetic glucocorticoid dexamethasone<sup>19,20</sup>. We found that dexamethasone effectively suppressed the restraint-induced rise in corticosterone to near basal levels in v-WT mice (Fig. 2d). However, the level of corticosterone in dexamethasone-injected v-TK mice was significantly higher than that in v-WT mice, consistent with a depressive-like phenotype.

We next examined whether adult neurogenesis regulates the behavioural response to stress in the novelty-suppressed feeding (NSF) test, which shows robust effects of antidepressants that are blocked by irradiation<sup>12,18</sup>. In this test, food-deprived mice are introduced to a novel open field containing a food pellet at its centre and the latency to begin feeding is recorded<sup>12</sup>. When assessed in the NSF test under normal conditions, v-WT and v-TK mice showed similar feeding latencies (Fig. 4a), indicating similar levels of approach-avoidance behaviour. Restraint stress just before testing, however, significantly increased the feeding latency in v-TK mice while having no effect on v-WT mice. Moreover, by the end of the 10 min test only 53% of stressed v-TK mice had fed compared with 92% of stressed v-WT mice (Fig. 4b). Mice from all groups consumed food upon returning to their home cage, indicating that decreased motivation to eat was not responsible for change in behaviour (Supplementary Fig. 6). Thus, adult neurogenesis does not alter behaviour under baseline conditions in this test, consistent with previous observations<sup>12,18</sup>, but buffers the effects of stress on feeding behaviour.

Because the NSF test is associated with both anxiety- and depressive-like behaviour, we investigated the interaction of neurogenesis and stress in additional behavioural tests. In the elevated plus maze, a test of anxiety-like behaviour, there was an anxiogenic effect of stress but no significant difference between v-WT and v-TK mice (Supplementary Fig. 7). We next tested depressive-like behaviour, using the forced swim test, in which rodents are placed in an inescapable cylinder of water and immobility is used as a measure of behavioural despair<sup>21</sup>. Under control conditions neurogenesis-deficient v-TK mice became immobile more rapidly and for a greater duration than v-WT mice (Fig. 4c, d), consistent with a depressive phenotype. In v-WT mice, restraint stress reduced the latency to become immobile and increased total immobility to the level of v-TK mice. Thus, neurogenesis-deficient mice displayed a depressive phenotype at baseline, which could be induced in intact mice by acute stress. Consistent with previous reports<sup>16,18,22</sup>, v-WT and v-TK mice showed similar levels of immobility during later stages of the test, when high (potentially ceiling) levels of behavioural despair are observed<sup>21</sup> (Supplementary Fig. 8).

Anhedonia is a hallmark symptom of depression and, in rodents, presents as a decrease in preference for a sucrose solution compared with water<sup>23</sup>. To assess whether adult neurogenesis is required for hedonic behaviours, we habituated v-WT and v-TK mice to freely available water and 1% sucrose for 3 days. Both groups similarly preferred sucrose (Supplementary Fig. 9a). Then, after water and sucrose deprivation, the bottles were reintroduced and preference was measured during a 10 min test. To introduce an aspect of reward-based decision making<sup>23</sup>, bottle locations were switched for the test. Although v-WT mice showed a preference for sucrose as before, neurogenesis-deficient v-TK mice showed no sucrose preference during the test (Fig. 4e). The decreased sucrose preference in v-TK mice was observed not only in the 10-min test but also during the next night (Fig. 4f). Overall consumption levels were not different, indicating that preference differences did not result from altered thirst or motivation to drink (Supplementary Fig. 9a–d). No differences were



**Figure 4 | Mice lacking neurogenesis show increased anxiety/depression-like behaviours.** **a**, In the NSF test, v-TK mice showed increased latency to feed in a novel environment after restraint stress but not under control conditions (genotype effect  $F_{1,75} = 5.9$ ,  $P = 0.02$ ; stress effect  $F_{1,75} = 1.9$ ,  $P = 0.17$ ; genotype  $\times$  stress interaction  $F_{1,75} = 4.8$ ,  $P = 0.03$ ;  $*P < 0.05$  versus v-TK control and  $*P < 0.01$  versus v-WT stressed;  $n = 13$ –25 per group). **b**, Cumulative distribution of feeding latencies for the NSF test (log-rank test;  $*P < 0.05$  versus all other groups). **c**, Neurogenesis-deficient v-TK mice became immobile faster in the forced swim test. Restraint stress reduced the latency to immobility in v-WT mice but did not affect v-TK mice (genotype effect  $F_{1,88} = 1.1$ ,  $P = 0.3$ ; stress effect  $F_{1,88} = 2.2$ ,  $P = 0.14$ ; genotype  $\times$  stress interaction  $F_{1,88} = 2.6$ ,  $P = 0.11$ ;  $*T_{46} = 2.1$ ,  $P < 0.05$  versus v-WT stressed; v-TK control versus v-TK stressed  $T_{42} = 0.1$ ,  $P = 0.9$ ;  $n = 22$ –26 per group). **d**, Under control conditions, the total time spent immobile was greater in v-TK mice than in v-WT mice. Restraint stress significantly increased total immobility in v-WT mice but had no effect on v-TK mice (genotype effect  $F_{1,88} = 0.3$ ,  $P = 0.6$ ; stress effect  $F_{1,88} = 4.2$ ,  $P = 0.04$ ; genotype  $\times$  stress interaction  $F_{1,88} = 9.1$ ,  $P = 0.003$ ;  $*P < 0.05$  versus control v-TK,  $*P < 0.001$  versus stressed v-WT, stressed v-WT versus stressed v-TK  $P > 0.05$ ;  $n = 22$ –26 per group). **e**, Neurogenesis-deficient v-TK mice showed reduced preference for sucrose in an acute test, compared with v-WT mice, under both control and restraint conditions (genotype effect  $F_{1,20} = 11.2$ ,  $P < 0.01$ ; stress effect  $F_{1,20} = 3.1$ ,  $P = 0.09$ ; genotype  $\times$  stress interaction  $F_{1,20} = 0.2$ ,  $P = 0.7$ ;  $n = 4$ –8 per group). **f**, Sucrose preference remained lower in v-TK mice than v-WT mice during the subsequent dark cycle (genotype effect  $F_{1,25} = 6.8$ ,  $P = 0.01$ ; stress effect  $F_{1,25} = 0.8$ ,  $P = 0.4$ ; genotype  $\times$  stress interaction  $F_{1,25} = 0.5$ ,  $P = 0.5$ ;  $n = 4$ –10 per group). Error bars, s.e.m.

observed in WT and control TK mice that were not treated with valganciclovir, indicating that anhedonic behaviour does not result from thymidine kinase expression alone (Supplementary Fig. 9e–g). The loss of sucrose preference in v-TK mice was observed whether or not the mice were restrained before testing, perhaps reflecting a basic difference between reward-related behaviours tested in this paradigm and the stress response behaviours tested in despair and avoidance situations. Taken together, our behavioural results suggest that adult



neurogenesis buffers the effect of previous stress in the NSF test, buffers the effect of inescapable stress in the forced swim test and enhances reward-seeking behaviour independent of stress in the sucrose preference test.

Elucidating the strong but poorly understood association between stress and depression is critical for development of more effective treatments<sup>1,2</sup>. The hippocampus has long been known to regulate the HPA axis<sup>6,7</sup>, and the importance of the hippocampus for emotional behaviour is emerging<sup>24</sup>. Our data show that adult-born hippocampal granule neurons dynamically regulate stress reactivity at both the endocrine and behavioural levels. A direct role for adult neurogenesis in depression-like behaviour was observed in three behavioural situations that are commonly used to assess antidepressant efficacy and characterize the development of a depressive phenotype in response to chronic stress<sup>12,18,25</sup>. In addition to its effects on emotional behaviour, stress is an important modulator of learning and memory<sup>26</sup>. Our results therefore also suggest that buffering of stress responses by new neurons may play a role in learning and memory under novel or aversive conditions, in addition to any more direct function of young neurons in encoding of information. Because the production of new granule neurons is itself strongly regulated by stress and glucocorticoids<sup>3</sup>, this system forms a loop through which stress, by inhibiting adult neurogenesis, could lead to enhanced responsiveness to future stress. This type of programming could be adaptive, predisposing animals to behave in ways best suited to the severity of their particular environments<sup>27</sup>. However, maladaptive progression of such a feed-forward loop could potentially lead to increased stress responsiveness and depressive behaviours that persist even in the absence of stressful events.

## METHODS SUMMARY

All procedures followed the Institute of Laboratory Animal Research guidelines and were approved by the Animal Care and Use Committee of the National Institute of Mental Health. Transgenic mice expressing HSV-TK under the GFAP promoter were generated from a previously generated plasmid<sup>28</sup> using standard techniques and bred on a mixed C57Bl/6:CD-1 background. Male v-WT and v-TK mice were treated with valganciclovir for 8 weeks (dexamethasone experiment), 10–19 weeks (endocrine), 12 weeks (behaviour) or 4 weeks (histology; histology after 12 weeks in Supplementary Fig. 1), beginning at 8 weeks of age. Male C57Bl/6 mice were irradiated under pentobarbital anaesthesia, as described previously<sup>29</sup>, and tested 9 weeks later. For immunohistochemical analyses, mice were given BrdU 6 weeks (for analysis of the paraventricular nucleus) or 24 h before being killed, brain sections were immunostained as previously described<sup>29</sup> and labelled cells were counted stereologically.

Serum corticosterone was measured by radioimmunoassay (MP Biomedicals) from submandibular blood samples obtained directly from the home cage condition or after exploration of a novel box, restraint or isoflurane exposure. For the dexamethasone suppression test, dexamethasone (Sigma; 50 µg kg<sup>-1</sup> in propylene glycol) or vehicle were injected 90 min before restraint, and blood was sampled immediately after 10 min restraint.

Behavioural tests were performed after 30 min of restraint or directly from the home cage. Different cohorts of mice were tested in the NSF test, elevated plus maze, forced swim test and sucrose preference test as previously described<sup>12,18,21,30</sup>. Statistical analyses were performed by *t*-test, log-rank test or ANOVA with Fisher's least significant difference test for post-hoc comparisons. Significance was set at *P* < 0.05.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 7 April; accepted 13 June 2011.

Published online 3 August 2011.

- Holsboer, F. & Ising, M. Stress hormone regulation: biological role and translation into therapy. *Annu Rev. Psychol.* **61**, 81–109 (2010).
- McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol. Rev.* **87**, 873–904 (2007).
- Mirescu, C. & Gould, E. Stress and adult neurogenesis. *Hippocampus* **16**, 233–238 (2006).

- Sapolsky, R. M. Is impaired neurogenesis relevant to the affective symptoms of depression? *Biol. Psychiatry* **56**, 137–139 (2004).
- Drew, M. R. & Hen, R. Adult hippocampal neurogenesis as target for the treatment of depression. *CNS Neurol. Disord. Drug Targets* **6**, 205–218 (2007).
- Roosendaal, B. et al. Memory retrieval impairment induced by hippocampal CA3 lesions is blocked by adrenocortical suppression. *Nature Neurosci.* **4**, 1169–1171 (2001).
- Jankord, R. & Herman, J. P. Limbic regulation of hypothalamo-pituitary-adrenocortical function during acute and chronic stress. *Ann. NY Acad. Sci.* **1148**, 64–73 (2008).
- Garcia, A. D., Doan, N. B., Imura, T., Bush, T. G. & Sofroniew, M. V. GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nature Neurosci.* **7**, 1233–1241 (2004).
- Seri, B., Garcia-Verdugo, J. M., McEwen, B. S. & Alvarez-Buylla, A. Astrocytes give rise to new neurons in the adult mammalian hippocampus. *J. Neurosci.* **21**, 7153–7160 (2001).
- Borrelli, E., Heyman, R. A., Arias, C., Sawchenko, P. E. & Evans, R. M. Transgenic mice with inducible dwarfism. *Nature* **339**, 538–541 (1989).
- Burns, K. A., Murphy, B., Danzer, S. C. & Kuan, C.-Y. Developmental and post-injury cortical gliogenesis: a genetic fate-mapping study with Nestin-CreER mice. *Glia* **57**, 1115–1129 (2009).
- Santarelli, L. et al. Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301**, 805–809 (2003).
- Furay, A. R., Bruestle, A. E. & Herman, J. P. The role of the forebrain glucocorticoid receptor in acute and chronic stress. *Endocrinology* **149**, 5482–5490 (2008).
- Wojtowicz, J. M. Irradiation as an experimental tool in studies of adult neurogenesis. *Hippocampus* **16**, 261–266 (2006).
- Doetsch, F., Caille, I., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716 (1999).
- Airan, R. D. et al. High-speed imaging reveals neurophysiological links to behavior in an animal model of depression. *Science* **317**, 819–823 (2007).
- Surget, A. et al. Antidepressants recruit new neurons to improve stress response regulation. *Mol. Psychiatry* doi:10.1038/mp.2011.48 (2011).
- David, D. J. et al. Neurogenesis-dependent and -independent effects of fluoxetine in an animal model of anxiety/depression. *Neuron* **62**, 479–493 (2009).
- Anacker, C., Zunszain, P. A., Carvalho, L. A. & Pariante, C. M. The glucocorticoid receptor: pivot of depression and of antidepressant treatment? *Psychoneuroendocrinology* **36**, 415–425 (2011).
- Ridder, S. et al. Mice with genetically altered glucocorticoid receptor expression show altered sensitivity for stress-induced depressive reactions. *J. Neurosci.* **25**, 6243–6250 (2005).
- Castagné, V., Moser, P. & Porsolt, R. D. in *Methods of Behavior Analysis in Neuroscience* 2nd edn (ed. Buccafusco, J. J.) Ch. 6 (CRC Press, 2009).
- Revest, J.-M. et al. Adult hippocampal neurogenesis is involved in anxiety-related behaviors. *Mol. Psychiatry* **14**, 959–967 (2009).
- Treadway, M. T. & Zald, D. H. Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neurosci. Biobehav. Rev.* **35**, 537–555 (2011).
- Barkus, C. et al. Hippocampal NMDA receptors and anxiety: at the interface between cognition and emotion. *Eur. J. Pharmacol.* **626**, 49–56 (2010).
- Duman, C. H. Models of depression. *Vitam. Horm.* **82**, 1–21 (2010).
- Sandi, C. & Pinelo-Nava, M. T. Stress and memory: behavioral effects and neurobiological mechanisms. *Neural Plast.* **2007**, 78970 (2007).
- Oomen, C. A. et al. Severe early life stress hampers spatial learning and neurogenesis, but improves hippocampal synaptic plasticity and emotional learning under high-stress conditions in adulthood. *J. Neurosci.* **30**, 6635–6645 (2010).
- Delaney, C. L., Brenner, M. & Messing, A. Conditional ablation of cerebellar astrocytes in postnatal transgenic mice. *J. Neurosci.* **16**, 6908–6918 (1996).
- Snyder, J. et al. Adult-born hippocampal neurons are more numerous, faster maturing, and more involved in behavior in rats than in mice. *J. Neurosci.* **29**, 14484–14495 (2009).
- Brown, T. S. & Murphy, H. M. Factors affecting sucrose preference behavior in rats with hippocampal lesions. *Physiol. Behav.* **11**, 833–844 (1973).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Sowers and J. Mitchell for help with irradiation, M. Brenner for providing the plasmid construct, GlaxoSmithKline for providing the antibody against HSV-TK, and K. Sanzone, F. Kamhi, L. Glover, S. Ferrante and L. Grigereit for assisting with mouse breeding and treatment. This research was supported by the Intramural Program of the National Institute of Mental Health, National Institutes of Health, USA, ZIA MH002784 (to H.A.C.).

**Author Contributions** J.S.S. performed histological, endocrine and behavioural experiments, analysed data and wrote the paper; A.S. performed behavioural experiments and analysed behavioural data; H.A.C. performed endocrine experiments, analysed data and wrote the paper; J.P. and M.B. generated the transgenic mice.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.A.C. ([heathercameron@mail.nih.gov](mailto:heathercameron@mail.nih.gov)).

## METHODS

All procedures followed the Institute of Laboratory Animal Research guidelines and were approved by the Animal Care and Use Committee of the National Institute of Mental Health.

**Transgenic animals.** Transgenic mice expressing HSV-TK under the GFAP promoter (GFAP-TK) were generated on a C57Bl/6 background using the pGfa2-TK1 plasmid<sup>28</sup> applying standard techniques and were outcrossed onto a mixed C57Bl/6:CD-1 background. For all GFAP-thymidine kinase experiments, male mice were weaned at 3 weeks of age, genotyped by PCR, and housed four per cage with mixed genotype siblings. Mice were housed on a 12 h light/dark cycle with lights on at 6.00 a.m. All experiments were performed on different, naive cohorts of mice except the home cage, exploration and 17 day restraint corticosterone experiments, which were performed on the same cohort. Beginning at 8 weeks of age, valganciclovir was mixed into powdered chow (0.3%, approximately 35 mg kg<sup>-1</sup> day<sup>-1</sup>) for 4 days, alternating with standard pelleted chow for 3 days. Mice were treated for 8 weeks before dexamethasone testing, 10–19 weeks before other endocrine testing and 12 weeks before behavioural testing. Histology was performed after 4 weeks, to assess effectiveness when animals were younger and neurogenesis was higher but was also routinely confirmed after longer treatments in tested mice (Supplementary Fig. 1). All mice within each experiment were tested at the same time point after the start of treatment.

**Irradiation.** Male C57Bl/6 mice were irradiated under pentobarbital anaesthesia, as described previously<sup>29</sup>, and tested 9 weeks later.

**Immunohistochemical analyses.** Immunohistochemistry was performed on 40- $\mu$ m sections as previously described<sup>29</sup>, using the following primary antibodies: goat anti-GFAP (Santa Cruz Biotechnology), rabbit anti-HSV-thymidine kinase (gift from GlaxoSmithKline), goat anti-doublecortin (Santa Cruz Biotechnology), rat anti-BrdU (Accurate) and mouse anti-BrdU (BD Biosciences). Alexa-conjugated secondary antibodies made in donkey (Invitrogen) and the nuclear counterstain Hoechst 33258 (Sigma) were used for all fluorescent labelling. The ABC method (Vector Labs) and DAB were used for enzymatic staining with cresyl violet nuclear counterstaining.

For dentate gyrus analyses, valganciclovir-treated and untreated WT and TK mice were injected once with BrdU (200 mg kg<sup>-1</sup>), after 4 weeks of treatment, and perfused 24 h later. Stereological counts of BrdU<sup>+</sup> cells were performed on a 1 in 12 series of sections in the dentate gyrus. Counts of BrdU<sup>+</sup>/DCX<sup>+</sup> neurons were obtained by multiplying total BrdU<sup>+</sup> cell counts by the proportion that expressed DCX. To quantify GFAP<sup>+</sup> astrocytes, the most medial 500  $\mu$ m of the molecular layer of the suprapyramidal blade was imaged with a confocal microscope (Olympus FV300) and a  $\times 60$  oil-immersion lens. The total number of GFAP<sup>+</sup> cells was divided by the volume examined for each animal to obtain a measure of astrocyte density.

For analyses of the paraventricular nucleus, mice were given BrdU water (1 mg ml<sup>-1</sup>) for 1 week, beginning in the fifth week of valganciclovir treatment, and perfused 6 weeks later. Stereological counts were performed in a one in two series of sections through the paraventricular nucleus.

For irradiation experiments, mice received a single BrdU injection and were perfused 24 h later. Immunostaining for BrdU and DCX was used to confirm a reduction in dentate gyrus neurogenesis, and BrdU and DCX labelling was examined in the subventricular zone of mice from the 30 min post-restraint group to assess the spatial specificity of irradiation-induced reduction of adult neurogenesis.

**Stress and corticosterone measurement.** To measure circadian fluctuations in baseline corticosterone, submandibular blood samples were obtained directly from the home cage condition, using animal lancets (Medipoint) at 7.00–8.00 a.m. or 7.00–8.00 p.m. Blood samples were centrifuged, and serum was collected and stored at  $-80^{\circ}\text{C}$  until use. Serum corticosterone was measured by radioimmunoassay (MP Biomedicals).

For exploration stress, mice were placed in an open field (white plastic box, 50 cm  $\times$  50 cm  $\times$  50 cm) for 5 min and allowed to explore. Blood samples were

obtained 30 min later. During the interval between exploration and blood sampling, mice were placed individually in clean, empty cages to prevent social interactions that could influence HPA activity. Morning and evening exploration occurred between 8.00 and 11.00 a.m. and 6.00 and 8.00 p.m., respectively. Locomotor behaviour was analysed using Ethovision software (Noldus).

For all restraint stress experiments, mice were restrained for 30 min in decapicones (Bainbridge Scientific). For isoflurane experiments, mice were exposed to 4% isoflurane in oxygen for 30 min. Blood samples were obtained 0, 30 or 60 min after stress on different subsets of mice; all mice were stressed at the same time and each mouse was used for only one of the three time points. For the 30- and 60-min groups, mice were individually placed in empty cages between stress and blood sampling. Restraint and isoflurane experiments were performed between 12.00 and 4.00 p.m.

For the dexamethasone suppression test, mice were injected with dexamethasone (Sigma; 50  $\mu\text{g kg}^{-1}$  in propylene glycol) or vehicle 90 min before restraint. Blood samples were taken immediately after 10 min restraint.

**Novelty-suppressed feeding.** Mice were handled for 2 min day<sup>-1</sup> for 3 days before testing to familiarize them with experimenter handling. Food was removed from the cage 24 h before testing. Mice were weighed just before food deprivation and again before testing to assess body weight loss. For testing, mice were placed for 10 min in a brightly lit open field (50 cm  $\times$  50 cm  $\times$  50 cm white plastic boxes containing bedding) with a food pellet at the centre on a slightly (1 cm) elevated platform. Mice were either placed in the arena directly from their home cage or after 30 min of restraint stress and 2 min back in their home cage. Behaviour was videotaped, and the latency for each mouse to begin feeding was scored, offline, by an experimenter blind to the genotype and condition for each mouse. Upon returning to their home cage, the total amount of food consumed during a 5-min period was analysed to test whether feeding differences in the novel environment were due to differences in hunger/motivation.

**Elevated plus maze.** Mice were handled for 3 days and were placed in an elevated plus maze for 5 min either directly from their home cage or after 30 min restraint and 2 min back in their home cage. Behaviour was tracked using Ethovision software and the amount of time spent in the open arms was measured.

**Forced swim test.** Mice were individually placed in a Plexiglas cylinder (19 cm diameter, 30 cm height) containing 19 cm water ( $23 \pm 1^{\circ}\text{C}$ ) and were videotaped for 6 min. Active (swimming, climbing, struggling) or passive (immobility) behaviours were scored using a time sampling technique to rate the predominant behaviour in each 5-s interval. In contrast to protocols designed to detect reductions in immobility<sup>21</sup> (for example, scoring minutes 2–6 of testing, when immobility is very high in controls, to detect antidepressant effects), the first 2 min of the test were scored separately to better detect potential increases in immobility. The latency to become immobile for the first time was also measured. After the swim session, mice were dried and placed in a cage surrounded by a heating pad. The water was changed between each animal.

**Sucrose preference.** Mice were not handled but were individually housed and given a water bottle containing water and a second with 1% sucrose with the left/right location balanced across animals. After 3 days of habituation, both bottles were removed at 12.00 p.m.<sup>30</sup>. Beginning at 6.30 p.m., mice were subjected to either 30-min restraint or brief experimenter handling, returned to their home cage, and given access to water and sucrose with bottles in the reversed left/right location. Bottles were weighed before testing and again after 10 min, and sucrose preference was expressed as  $(\Delta\text{weight}_{\text{sucrose}})/(\Delta\text{weight}_{\text{sucrose}} + \Delta\text{weight}_{\text{water}}) \times 100$ . The bottles were weighed again at 1 h and on the subsequent morning, and sucrose preference over this interval was used as a long-term, overnight measure.

**Statistical analyses.** Statistical analyses were performed by *t*-test, log-rank test or ANOVA with Fisher's LSD test for post-hoc comparisons. Significance was set at  $P < 0.05$ .

# Southern Ocean dust–climate coupling over the past four million years

Alfredo Martínez-García<sup>1,2,3</sup>, Antoni Rosell-Melé<sup>3,4,5</sup>, Samuel L. Jaccard<sup>1</sup>, Walter Geibert<sup>6,7</sup>, Daniel M. Sigman<sup>8</sup> & Gerald H. Haug<sup>1,2</sup>

Dust has the potential to modify global climate by influencing the radiative balance of the atmosphere and by supplying iron and other essential limiting micronutrients to the ocean<sup>1,2</sup>. Indeed, dust supply to the Southern Ocean increases during ice ages, and ‘iron fertilization’ of the subantarctic zone may have contributed up to 40 parts per million by volume (p.p.m.v.) of the decrease (80–100 p.p.m.v.) in atmospheric carbon dioxide observed during late Pleistocene glacial cycles<sup>3–7</sup>. So far, however, the magnitude of Southern Ocean dust deposition in earlier times and its role in the development and evolution of Pleistocene glacial cycles have remained unclear. Here we report a high-resolution record of dust and iron supply to the Southern Ocean over the past four million years, derived from the analysis of marine sediments from ODP Site 1090, located in the Atlantic sector of the subantarctic zone. The close correspondence of our dust and iron deposition records with Antarctic ice core reconstructions of dust flux covering the past 800,000 years (refs 8, 9) indicates that both of these archives record large-scale deposition changes that should apply to most of the Southern Ocean, validating previous interpretations of the ice core data. The extension of the record beyond the interval covered by the Antarctic ice cores reveals that, in contrast to the relatively gradual intensification of glacial cycles over the past three million years, Southern Ocean dust and iron flux rose sharply at the Mid-Pleistocene climatic transition around 1.25 million years ago. This finding complements previous observations over late Pleistocene glacial cycles<sup>5,8,9</sup>, providing new evidence of a tight connection between high dust input to the Southern Ocean and the emergence of the deep glaciations that characterize the past one million years of Earth history.

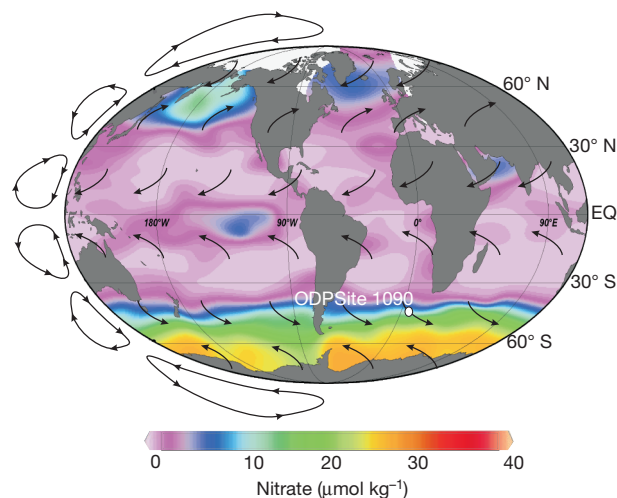
The scarcity of iron reduces marine export production and carbon uptake in one-quarter of the world ocean where the concentration of major nutrients (phosphorus and nitrogen) is perennially high<sup>10</sup> (Fig. 1). The Southern Ocean is the region where variations in iron availability can have the largest effect on Earth’s carbon cycle through its fertilizing effect on marine ecosystems.

In the modern Southern Ocean, deep ocean waters with high CO<sub>2</sub> and nutrient content are brought to the surface by wind-driven upwelling and density-driven overturning. However, the scarcity of iron reduces phytoplankton growth<sup>1,10</sup>, and major nutrients are returned to the subsurface before they are fully consumed. This incomplete utilization of nutrients represents an important leak in the modern global biological pump because it allows the escape of deeply sequestered carbon back to the atmosphere, thereby raising atmospheric CO<sub>2</sub> levels<sup>6</sup>. Today, the Antarctic zone is the greater part of the Southern Ocean CO<sub>2</sub> leak, because of its extremely low degree of surface nutrient consumption and its ventilation of the abyssal ocean. The subantarctic zone, further north in the Southern Ocean, is characterized by greater nutrient consumption and ventilates the somewhat smaller volume of the mid-depth ocean. However, its nutrient status affects the productivity of the low

latitude ocean, with repercussions for low-latitude CaCO<sub>3</sub> production and for the partitioning of regenerated carbon storage between the mid-depth and deep ocean. Thus, both the Antarctic zone and the subantarctic zone are potentially important for glacial/interglacial CO<sub>2</sub> change<sup>6,7</sup>.

We first focus on the Antarctic zone. An increase in the efficiency of the global biological pump can be accomplished either by decreasing the physical cycling of deep water through the surface of the Antarctic zone<sup>11</sup> or by increasing the degree to which Antarctic zone surface nutrients are consumed by marine organisms (for example, as a result of an increase in the availability of iron relative to other nutrients<sup>6,12</sup>). Data and models have been used to argue that the Antarctic was more strongly stratified during ice ages, reducing the Antarctic leak in the biological pump<sup>6,12</sup>. Because Antarctic major nutrient and iron supply come from upwelled water, productivity in the Antarctic zone would probably decline during ice ages in step with stratification, consistent with most Antarctic productivity data<sup>4</sup>. Nevertheless, the greater the input of aeolian iron, the greater the degree of nutrient consumption is expected for a given increase in Antarctic stratification, because the dust input raises the total iron supply to the surface ocean relative to the input of major nutrients and excess CO<sub>2</sub> from below<sup>6,7</sup>. Thus, dust input to the Antarctic is a central regulator of how completely the Antarctic CO<sub>2</sub> leak can be sealed during ice ages, regardless of whether productivity increased or decreased in the region.

Second, we consider the subantarctic zone. This zone, unlike the Antarctic, is characterized by higher productivity during ice ages<sup>4,5,13</sup>.



**Figure 1 | Location of ODP Site 1090, world ocean surface nitrate concentrations, and wind direction.** ODP Site 1090, the source of the sediment cores used in this study, is located at 42° 54.5' S, 8° 54.0' E, at 3,702 m depth. Nitrate concentrations are from the Electronic Atlas of the World Ocean Experiment. Black arrows are schematic representations of atmospheric convection cells and wind directions. EQ, Equator.

<sup>1</sup>Geological Institute, ETH Zürich, Zürich 8092, Switzerland. <sup>2</sup>DFG-Leibniz Center for Surface Process and Climate Studies, Institute for Geosciences, Potsdam University, Potsdam D-14476, Germany. <sup>3</sup>Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona, Bellaterra 08193, Catalonia, Spain. <sup>4</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Catalonia, Spain. <sup>5</sup>College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331-5503, USA. <sup>6</sup>School of GeoSciences, The University of Edinburgh, Edinburgh EH9 3JW, UK. <sup>7</sup>Scottish Association for Marine Science (SAMS), Scottish Marine Laboratory, Oban, Argyll PA37 1QA, UK. <sup>8</sup>Department of Geosciences, Princeton University, New Jersey 08544, USA.



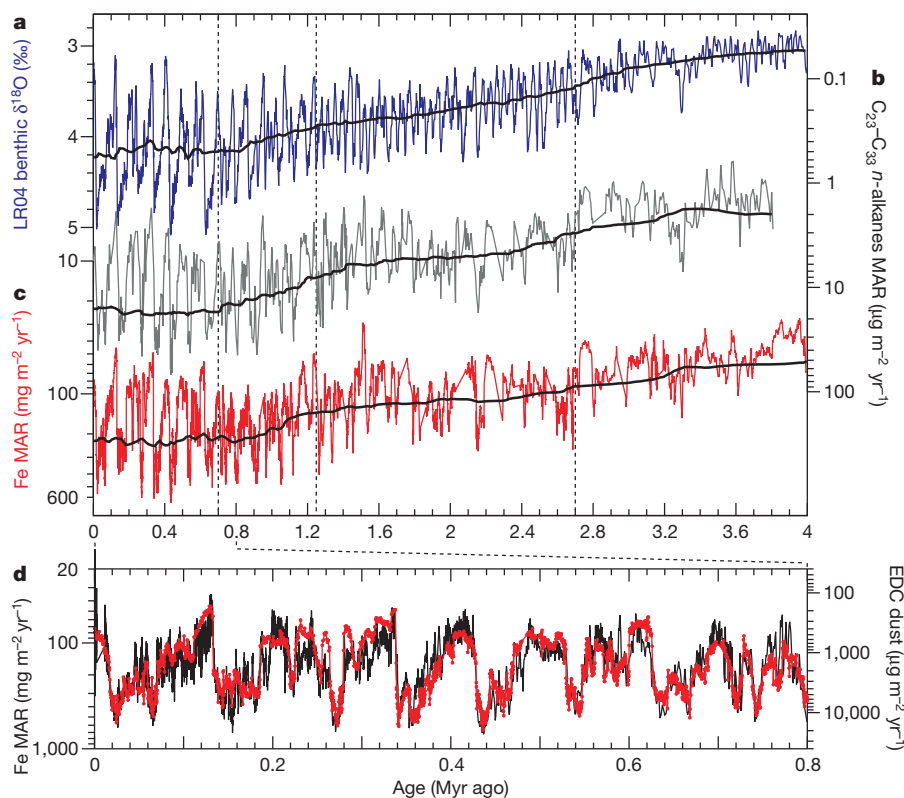
This suggests a conceptually simpler mechanism than in the Antarctic zone by which dust-driven iron fertilization of the subantarctic zone may have lowered atmospheric CO<sub>2</sub>, with the increase in subantarctic zone productivity consuming a greater fraction of the surface nutrients and thus driving more storage of carbon in the ocean interior. The subantarctic zone is one of the best candidates for aeolian iron fertilization, because it falls in the same latitude as the major Southern Hemisphere dust sources, and is on the wind path from them<sup>3,6</sup>. Further, there is evidence that natural aeolian iron deposition enhances marine export production in the modern subantarctic zone<sup>14</sup>. Palaeoceanographic data appear to constrain its effect to a maximum of 40 p.p.m.v. of the atmospheric CO<sub>2</sub> decrease observed in the last glacial cycle, because a significant increase in dust fluxes and marine export production in the subantarctic zone is only observed in the transition from marine isotope stage (MIS) 5 to MIS 4, when atmospheric CO<sub>2</sub> concentrations are already around 40–50 p.p.m.v. lower than during the interglacial periods<sup>3–7</sup>. A CO<sub>2</sub> drawdown of ≤40 p.p.m. from subantarctic zone iron fertilization is consistent with estimates obtained using geochemical box models and Earth system models, which in most cases range between 20 and 40 p.p.m.v. (see refs 3, 7, 15–18 and Supplementary Information).

The ice core reconstruction generated by the EPICA project<sup>8,9</sup> has provided a unique record with which to test dust–climate interactions in the Southern Ocean over the past 800,000 years (Fig. 2d). However, ice core records only allow an indirect inference of the supply of dust and iron to the ocean, and it is unclear how far back in time they can be extended<sup>19</sup>. Here we report a high-resolution dust and iron record from the subantarctic Atlantic spanning the past four million years (4 Myr), which provides the first insights into the variability of Southern Ocean dust deposition through the major climatic transitions of the Pliocene and Pleistocene epochs.

The record is based on the combined analysis of independent organic (*n*-alkanes) and inorganic (Fe, Ti and <sup>232</sup>Th) dust proxies in

marine sediments from ODP Site 1090 (Fig. 1). Long-chain *n*-alkanes are lipid constituents of the epicuticular wax layer of terrestrial plants. These terrestrial lipids are eroded from leaf surfaces and soils by winds and transported in the organic fraction of aeolian dust over the ocean<sup>20</sup>. In our record, there is a clear predominance of long-chain *n*-alkanes with odd carbon numbers over those with even carbon numbers, over the Pliocene and Pleistocene epochs. This is representative of inputs of leaf waxes from terrestrial plants transported by wind<sup>20</sup>, supporting the interpretation of the record in terms of changes in dust deposition through time<sup>5</sup>. Fe, Ti and <sup>232</sup>Th are elements with a well-characterized lithogenic origin. Hence, their concentration in marine sediments can be used to quantitatively estimate variations in the supply of wind-borne lithogenic material to remote areas of the ocean where other sources of continental material are negligible<sup>5,13,21</sup> (see Supplementary Information). The consistency among the different dust tracers measured in ODP Site 1090 and their excellent correlation with the dust deposition record from the Antarctic ice cores over the past 0.8 Myr (Fig. 2d) indicate that they are faithful recorders of changes in the supply of dust, and hence of iron, to high southern latitudes through time. This subantarctic zone/ice-core agreement in dust flux history also validates the long-held assumption that the ice core records reflect the pattern of dust fluxes to the Southern Ocean surface. A difference in the magnitude of the glacial–interglacial change of dust flux between ocean sediment and polar ice core records has been noted previously<sup>5,8</sup>. We note a similar difference here, with our record indicating a smaller glacial–interglacial change compared with the changes of more than an order of magnitude indicated by the Antarctic ice cores (Fig. 2d). This difference is probably due to variations in the efficiency of dust transport mechanisms to the Antarctic continent during glacial stages<sup>8</sup>.

Our record reveals that, on orbital timescales, Southern Ocean dust deposition has been exponentially linked to global ice volume changes through the Pliocene and Pleistocene epochs (Fig. 2). Major steps in



**Figure 2 | Global ice volume, dust and iron deposition in the Southern Ocean since 4 Myr ago.** **a**, Lisiecki and Raymo (LR04) benthic  $\delta^{18}\text{O}$  stack<sup>28</sup>. **b**, **c**, Mass accumulation rates (MARs) of *n*-alkanes (**b**) and iron (**c**) at ODP Site 1090. Both records are plotted on a reversed logarithmic scale to emphasize the

exponential relationship with the benthic  $\delta^{18}\text{O}$  record. **d**, Expanded view of iron variability over the past 800 kyr at ODP Site 1090 (red line), plotted together with dust flux estimates in the Antarctic ice core from EPICA Dome C (EDC; black line)<sup>8</sup>.

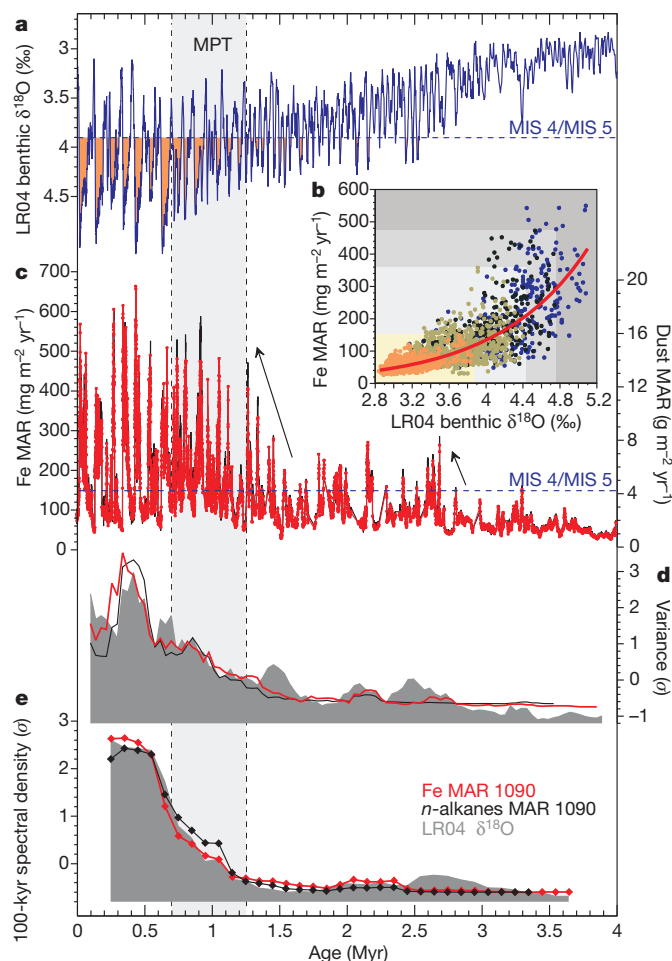
dust deposition coincide with the two major climatic transitions of the Pliocene–Pleistocene, indicating a pervasive link between Southern Ocean dust and Northern Hemisphere glaciations.

The first significant rise in ice-age dust and iron deposition is observed around 2.7 Myr ago, coinciding with the initiation of major glaciations in the Northern Hemisphere (Figs 1 and 2). The development of large, perennial Northern Hemisphere ice sheets around 2.7 Myr ago represents a milestone in climate evolution that marks the end of the relatively stable Pliocene warm period and the emergence of strong feedbacks in the climate system that characterize Quaternary climate oscillations. Although Earth's orbital configuration favoured the initiation of regional glaciations, a decrease in atmospheric  $\text{CO}_2$  concentrations may be required to initiate and sustain larger glaciations<sup>22</sup>. Data from the Antarctic zone and the subarctic North Pacific suggest that a transition towards permanent polar ocean water-column stratification identified around 2.7 Myr ago may have been a key driver of this  $\text{CO}_2$  decline<sup>23</sup>. Our data indicate that Antarctic zone stratification may have been reinforced by an increase in iron supply to the Southern Ocean during glacial stages after 2.7 Myr ago, and a similar dynamic may have applied to the subarctic North Pacific.

During the Pliocene, meridional temperature gradients were relatively weak, and the tropical warm pool was apparently more expansive than today<sup>24,25</sup>. This climate state resulted in relatively weak atmospheric convective cells that extended further poleward, greatly affecting wind intensity and global precipitation patterns<sup>24</sup>. As a result of these conditions, models predict stronger precipitation over Africa, Australia and South America during Pliocene times<sup>24</sup>, which, together with a weak atmospheric circulation, would have minimized dust generation and transport in these regions. Thus, the strengthening of the meridional temperature gradients at the 2.7-Myr transition and the mid-Pleistocene transition (MPT) as a consequence of the increase in ice volume would have resulted in intensification and spatial contraction of the atmospheric convective cells<sup>24,25</sup> and a reduction in precipitation; these effects would have favoured dust generation and transport, consistent with the increases in dust and iron flux observed at Site 1090.

The most dramatic rise in dust deposition in our record occurs across the MPT (Fig. 3). During this period, Southern Ocean dust fluxes doubled, reaching peak fluxes comparable to those of the Last Glacial Maximum. Although the controls on Southern Ocean dust deposition are not well understood, previous studies have suggested that dust generation requires that a climate threshold must be reached before the system is sensitive to perturbation<sup>5,8,18</sup>. Taking the last glacial cycle as an example, dust flux changed relatively little during the initial climate deterioration during MIS 5, increasing abruptly at MIS 4 with the rapid increase in Northern Hemisphere ice<sup>5,7</sup>. Our long-term dust flux and iron deposition records are strongly convergent with this view, and reveal that Southern Ocean dust flux has a unique sensitivity to the very cold glacial conditions that characterize the past 1.25 Myr. Whereas measures of climate cooling and glaciation indicate a relatively gradual trend in mean climate over the Pleistocene, dust flux at Site 1090 increases exponentially at the MPT, when the most intense ice ages first appear (Fig. 3). The data thus confirm and extend the evidence for a nonlinear relationship of Southern Ocean dust flux to climate (Fig. 3b), indicating that it applies to the entire Pliocene–Pleistocene. This sensitivity of dust flux to climate is also consistent with the observations of relatively constant interglacial tropical sea surface temperature and atmospheric  $\text{CO}_2$  across the MPT<sup>26,27</sup>, as no significant changes in interglacial dust deposition are observed.

Beginning at 1.25 Myr ago, the MPT was characterized by the development of thicker and less stable ice-sheets in the Northern Hemisphere and the emergence of the high-amplitude, asymmetric 100-kyr cycle that distinguishes late Pleistocene glaciations<sup>28,29</sup>. This important climatic shift occurred in the absence of any significant change in orbital forcing, leading to the view that the MPT involved a reorganization of the climate system's internal feedbacks<sup>29</sup>. A



**Figure 3 | Evolution of global ice volume, and Southern Ocean dust and iron variability through the Pliocene and Pleistocene epochs.** **a**, Benthic  $\delta^{18}\text{O}$  stack<sup>28</sup>. Orange shading indicates the intervals where glaciations intensified and dust fluxes rose to levels comparable to those of the Marine Isotope Stage (MIS) 5 to MIS 4 transition. **b**, Regression between benthic  $\delta^{18}\text{O}$  and Fe MAR. Colours indicate key time intervals: 0–0.7 Myr ago (blue), 0.7–1.25 Myr ago (black), 1.25–2.7 Myr ago (brown) and 2.7–4 Myr ago (orange). Background shading areas highlight the evolution of the Fe MAR/benthic  $\delta^{18}\text{O}$  coupling across the defined key time intervals. **c**, Fe MAR (red line), and dust MAR estimated from the Ti MAR (black line). **d**, Normalized variance of Fe MAR, *n*-alkanes MAR, and benthic  $\delta^{18}\text{O}$ , computed in 200-kyr windows with a number of lags equal to half the timeseries length (' $\frac{1}{2}$  lags'); key given in **e**. **e**, Normalized power spectral density of the 100-kyr cycle in Fe MAR, *n*-alkanes MAR and  $\delta^{18}\text{O}$  computed using 500-kyr windows and  $\frac{1}{2}$  lags. Vertical dashed lines with grey shading between indicate start and finish of MPT according to ref. 29. In **d** and **e**, the variance and the power spectral density of the 100-kyr cycle of the different records were normalized to zero mean and unit standard deviation ( $\sigma$ ).

decrease in atmospheric  $\text{CO}_2$  concentrations and the associated global cooling is often proposed as one of the mechanisms to explain the MPT<sup>29</sup>. The available data do indeed suggest that atmospheric  $\text{CO}_2$  concentrations during peak ice age conditions decreased by  $\sim 30$  p.p.m. across the MPT<sup>26</sup>.

Although our data cannot yet enable us to prove causation, they raise the strong possibility that dust not only increased into the stronger ice ages of the post-MPT world but then also acted as a positive feedback, encouraging further  $\text{CO}_2$  reduction and cooling through iron fertilization of the Southern Ocean. Consistent with this hypothesis, an increase in opal mass accumulation rate (MAR) has also been observed in the Atlantic sector of the Southern Ocean during this period, coinciding with the development of extensive diatom mats between  $47^\circ\text{S}$  and  $50^\circ\text{S}$  and with the emergence of the diatom *Fragilariopsis kerguelensis*, the dominant species in modern Southern Ocean sediments<sup>30</sup>. These

observations suggest that an increase in iron availability may have driven part or all of the 30 p.p.m.v. decrease in peak glacial atmospheric CO<sub>2</sub> observed across the MPT. By driving the descent into deep, cold glacial periods, the climate/dust/CO<sub>2</sub> feedback may underpin the development of the strong 100-kyr periodicity that characterizes global climate and dust since the MPT (Fig. 3e). In detail, the greater ice-age dust loads may have increased carbon sequestration by both the Antarctic zone—where it would have encouraged increasingly complete major nutrient consumption even if circulation changes lowered total ice age productivity—and the subantarctic zone, where ice age iron fertilization appears to have fuelled a simple increase in export production.

## METHODS SUMMARY

The methods used in the analysis of *n*-alkanes, Fe, Ti and <sup>232</sup>Th have been described in detail elsewhere<sup>5</sup>. The high-resolution records of Fe and Ti measured with the X-ray fluorescence (XRF) core-scanner were converted into quantitative concentrations using more than 400 inductively coupled plasma sector-field mass spectrometry (ICP-SFMS) measurements (Supplementary Fig. 1). Dry bulk density (DBD) was estimated for each XRF measurement interval from the  $\gamma$ -ray attenuation (GRA) density measurements performed in the same sediment sections as the XRF. MAR was calculated by multiplying the concentration of the different compounds by the DBD and the sedimentation rate. After error propagation, we find that the analytical component of the uncertainty is 7.8%, 8.4% and 8.6% of the final value (1 $\sigma$ ) for Fe, Ti and *n*-alkanes MAR, respectively. The influence of age model uncertainty on the estimated MAR is evaluated in Supplementary Fig. 6. Our results show that the analytical uncertainty and the potential errors in the age model do not result in significant deviations from the estimated MAR that would affect the conclusions of this study.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 14 February; accepted 16 June 2011.**

**Published online 3 August 2011.**

- Martin, J. H., Gordon, R. M. & Fitzwater, S. E. Iron in Antarctic waters. *Nature* **345**, 156–158 (1990).
- Martin, J. Glacial-interglacial CO<sub>2</sub> change: the iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Watson, A. J., Bakker, D. C. E., Ridgwell, A. J., Boyd, P. W. & Law, C. S. Effect of iron supply on Southern Ocean CO<sub>2</sub> uptake and implications for glacial atmospheric CO<sub>2</sub>. *Nature* **407**, 730–733 (2000).
- Kohfeld, K. E., Le Quere, C., Harrison, S. P. & Anderson, R. F. Role of marine biology in glacial-interglacial CO<sub>2</sub> cycles. *Science* **308**, 74–78 (2005).
- Martinez-Garcia, A. *et al.* Links between iron supply, marine productivity, sea surface temperature, and CO<sub>2</sub> over the last 1.1 Ma. *Paleoceanography* **24**, PA1207, doi:10.1029/2008PA001657 (2009).
- Hain, M. P., Sigman, D. M. & Haug, G. H. The polar ocean and glacial cycles in atmospheric CO<sub>2</sub> concentration. *Nature* **466**, 47–55 (2010).
- Hain, M. P., Sigman, D. M. & Haug, G. H. Carbon dioxide effects of Antarctic stratification, North Atlantic Intermediate Water formation, and subantarctic nutrient drawdown during the last ice age: diagnosis and synthesis in a geochemical box model. *Glob. Biogeochem. Cycles* **24**, GB4023, doi:10.1029/2010gb003790 (2010).
- Lambert, F. *et al.* Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**, 616–619 (2008).
- Wolff, E. W. *et al.* Southern Ocean sea-ice extent, productivity and iron flux over the past eight glacial cycles. *Nature* **440**, 491–496 (2006).
- Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science* **315**, 612–617 (2007).
- Toggweiler, J. R. Variation of atmospheric CO<sub>2</sub> by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
- François, R. *et al.* Contribution of Southern Ocean surface-water stratification to low atmospheric CO<sub>2</sub> concentrations during the last glacial period. *Nature* **389**, 929–935 (1997).
- Kumar, N. *et al.* Increased biological productivity and export production in the glacial Southern Ocean. *Nature* **378**, 675–680 (1995).
- Cassar, N. *et al.* The Southern Ocean biological response to aeolian iron deposition. *Science* **317**, 1067–1070 (2007).
- Matsumoto, K., Sarmiento, J. L. & Brzezinski, M. A. Silicic acid leakage from the Southern Ocean: a possible explanation for glacial atmospheric pCO<sub>2</sub>. *Glob. Biogeochem. Cycles* **16**, 1031, doi:10.1029/2001GB001442 (2002).
- Brovkin, V., Ganopolski, A., Archer, D. & Rahmstorf, S. Lowering of glacial atmospheric CO<sub>2</sub> in response to changes in oceanic circulation and marine biogeochemistry. *Paleoceanography* **22**, PA4202, doi:10.1029/2006PA001380 (2007).
- Kohler, P., Fischer, H., Munhoven, G. & Zeebe, R. E. Quantitative interpretation of atmospheric carbon records over the last glacial termination. *Global Biogeochem. Cycles* **19**, GB4020, doi:10.1029/2004GB002345 (2005).
- Ridgwell, A. J. Implications of the glacial CO<sub>2</sub> “iron hypothesis” for Quaternary climate change. *Geochim. Geophys. Geosyst.* **4**, 1076, doi:10.1029/2003GC000563 (2003).
- Jouzel, J. & Masson-Delmotte, V. Deep ice cores: the need for going back in time. *Quat. Sci. Rev.* **29**, 3683–3689 (2010).
- Simoneit, B. R. T., Chester, R. & Eglinton, G. Biogenic lipids in particulates from the lower atmosphere over the eastern Atlantic. *Nature* **267**, 682–685 (1977).
- Winckler, G., Anderson, R. F., Fleisher, M. Q., McGee, D. & Mahowald, N. Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. *Science* **320**, 93–96 (2008).
- Lunt, D. J., Foster, G. L., Haywood, A. M. & Stone, E. J. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO<sub>2</sub> levels. *Nature* **454**, 1101–1105 (2008).
- Sigman, D. M., Jaccard, S. L. & Haug, G. H. Polar ocean stratification in a cold climate. *Nature* **428**, 59–63 (2004).
- Brierley, C. M. *et al.* Greatly expanded tropical warm pool and weakened Hadley Circulation in the early Pliocene. *Science* **323**, 1714–1718 (2009).
- Martinez-Garcia, A., Rosell-Mele, A., McClymont, E. L., Gersonde, R. & Haug, G. H. Subpolar link to the emergence of the modern equatorial Pacific Cold Tongue. *Science* **328**, 1550–1553 (2010).
- Honisch, B., Hemming, N. G., Archer, D., Siddall, M. & McManus, J. F. Atmospheric carbon dioxide concentration across the mid-Pleistocene transition. *Science* **324**, 1551–1554 (2009).
- Herbert, T. D., Peterson, L. C., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328**, 1530–1534 (2010).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003, doi:10.1029/2004PA001071 (2005).
- Clark, P. U. *et al.* The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO<sub>2</sub>. *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
- Cortese, G. & Gersonde, R. Plio/Pleistocene changes in the main biogenic silica carrier in the Southern Ocean, Atlantic Sector. *Mar. Geol.* **252**, 100–110 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Stefer for performing the XRF scanner measurements at the University of Bremen; and I. Vöge for assistance in the ICP-SFMS analysis at the Alfred Wegener Institute for Polar and Marine Research. We thank the Integrated Ocean Drilling Program for providing the samples used in this study. This research used data acquired at the XRF Core Scanner Laboratory at the MARUM – Center for Marine Environmental Sciences, University of Bremen. Support for this work was provided by the Spanish Ministry of Science and Innovation (MPCINN), the European Commission, and the Deutsche Forschungsgemeinschaft (DFG).

**Author Contributions** A.M.-G., A.R.-M. and G.H.H. designed the study. A.M.-G. performed the *n*-alkane and elemental ICP-SFMS analysis and wrote the first version of the manuscript. G.H.H. and S.L.J. organized and supervised the XRF scanning at the University of Bremen. W.G. organized and supervised the ICP-SFMS elemental analysis at the Alfred Wegener Institute. All the authors contributed to the interpretation of the data and provided significant input to the final manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.M.-G. ([alfredo.martinez-garcia@erdw.ethz.ch](mailto:alfredo.martinez-garcia@erdw.ethz.ch)).



## METHODS

**Analysis of *n*-alkanes.** The method used for analysis of *n*-alkanes has been described in detail elsewhere<sup>5,31,32</sup>. Briefly, sediment samples were freeze-dried, homogenized, microwave extracted, and analysed with a Thermo Trace GC-FID. The identification of the different compounds was achieved through comparison of the chromatographic relative retention times with those of the target compounds in standards. Several selected samples were analysed by mass spectrometry to confirm peak identities and the absence of co-eluting peaks. The reproducibility of the procedure was evaluated using a homogeneous sediment standard, extracted with every batch of 14 samples. The external analytical reproducibility in the determination of *n*-alkanes in the different batches analysed was 7%.

**ICP-SFMS analysis of Fe, Ti, <sup>232</sup>Th.** The methods used in the analysis of Fe, Ti and <sup>232</sup>Th have been described in detail elsewhere<sup>5</sup>. Briefly, sediment samples were freeze-dried and digested in a pressure-assisted microwave system using an acid mixture of HNO<sub>3</sub>, HCl and HF. Fe, Ti and <sup>232</sup>Th concentrations were determined by inductively coupled plasma sector-field mass spectrometry (ICP-SFMS, Element2, Thermo Scientific). The calibration was done with dilutions of a standard solution, and Rh was used as internal standard. External reproducibility was evaluated using the NIST standard reference material 2702 ("inorganics in marine sediment"), and with each batch of samples, two procedural blanks were run. The average relative error for Fe and Ti was respectively 1.8% and 2.1% (1 $\sigma$ ), and the external reproducibility of NIST 2702 was 5.6% and 6.5% respectively, for the batches included in Supplementary Fig. 1.

**XRF scanner measurements.** The high resolution elemental analysis of Fe and Ti was performed using an Aavatech profiling X-ray Fluorescence (XRF) core scanner at Bremen University at a 1 cm downcore resolution. The external reproducibility of the Aavatech XRF core scanner for Fe and Ti in the range of the measurements is below 2% (1 $\sigma$ ).

**Age model.** In the interval 0–800 kyr ago, we use the age model generated by graphic correlation of the XRF Fe measurements to the ice core dust reconstruction from the EPICA project<sup>8</sup> (Supplementary Fig. 2). This allows us to compare the marine and continental records on similar timescales. In the interval 800–2,900 kyr ago, we use the original age model of ODP Site 1090 based on benthic  $\delta^{18}\text{O}$  stratigraphy<sup>33</sup>. In the older part of the record we have slightly modified the age model based on biostratigraphy<sup>34,35</sup> by aligning the *n*-alkane record to the Lisiecki and Raymo (LR04) global benthic  $\delta^{18}\text{O}$  stack (see Supplementary Fig. 3).

**Mass accumulation rate (MAR) calculation.** The high-resolution estimates of Fe and Ti measured with the XRF core-scanner have been converted into quantitative

concentrations using more than 400 ICP-SFMS measurements. The correlation coefficients and the regression lines used for the calculation are shown in Supplementary Fig. 1. Dry bulk density (DBD) has been estimated for each XRF measurement interval from the  $\gamma$ -ray attenuation (GRA) density measurements<sup>35</sup> performed in the same sediment sections as the XRF. GRA bulk density has been converted into DBD using the regression line shown in Supplementary Fig. 1a, which is obtained from the cross-correlation of the GRA density and the available DBD discrete measurements<sup>35</sup>. The errors associated with the DBD are less 5% of the estimated value<sup>36</sup>. MARs were calculated by multiplying the concentration of the different compounds by the DBD and the sedimentation rate.

**MAR uncertainty.** In order to assess the uncertainty in the MAR of the different elements, we consider the errors of the different measurements and how they are propagated in the calculation. In order to do so, we distinguish between statistical errors, derived from the uncertainty associated with the analysis, and systematic deviations, as potentially present in the age model. Systematic deviations of the age model cannot be treated like statistical errors, as they do not meet the requirements for the underlying mathematics. After error propagation, we find that the analytical component of the uncertainty is 7.8%, 8.4% and 8.6% of the final value (1 $\sigma$ ) for Fe, Ti and *n*-alkanes MAR respectively. The influence of age model uncertainty on the estimated MAR is evaluated in Supplementary Fig. 6. Maximum age model error envelopes of 4 kyr, 6 kyr and 15 kyr for the intervals 0–1 Myr ago, 1–3 Myr ago and 3–4 Myr ago respectively<sup>28</sup> do not allow for significant changes in the estimated MAR that would affect the conclusions of this study.

31. McClymont, E. L., Martinez-Garcia, A. & Rosell-Mele, A. Benefits of freeze-drying sediments for the analysis of total chlorins and alkenone concentrations in marine sediments. *Org. Geochem.* **38**, 1002–1007 (2007).
32. Kornilova, O. & Rosell-Mele, A. Application of microwave-assisted extraction to the analysis of biomarker climate proxies in marine sediments. *Org. Geochem.* **34**, 1517–1523 (2003).
33. Venz, K. A. & Hodell, D. A. New evidence for changes in Plio-Pleistocene deep water circulation from Southern Ocean ODP Leg 177 Site 1090. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **182**, 197–220 (2002).
34. Zielinski, U. & Gersonde, R. Plio-Pleistocene diatom biostratigraphy from ODP Leg 177, Atlantic sector of the Southern Ocean. *Mar. Micropaleontol.* **45**, 225–268 (2002).
35. Gersonde, R. et al. Southern Ocean paleoceanography. Sites 1088–1094. *Proc. ODP Init. Rep.* **177**, (1999).
36. Blum, P. *Physical Properties Handbook: A Guide to the Shipboard Measurement of Physical Properties of Deep-Sea Cores* (ODP Tech. Note 26, 1997).

# RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia

Johannes Zuber<sup>1,2\*</sup>, Junwei Shi<sup>1,3\*</sup>, Eric Wang<sup>1</sup>, Amy R. Rappaport<sup>1,4</sup>, Harald Herrmann<sup>5</sup>, Edward A. Sison<sup>6</sup>, Daniel Magoon<sup>6</sup>, Jun Qi<sup>7</sup>, Katharina Blatt<sup>8</sup>, Mark Wunderlich<sup>9</sup>, Meredith J. Taylor<sup>1</sup>, Christopher Johns<sup>1</sup>, Agustin Chicas<sup>1</sup>, James C. Mulloy<sup>9</sup>, Scott C. Kogan<sup>10</sup>, Patrick Brown<sup>6</sup>, Peter Valent<sup>5,8</sup>, James E. Bradner<sup>7</sup>, Scott W. Lowe<sup>1,4,11</sup> & Christopher R. Vakoc<sup>1</sup>

Epigenetic pathways can regulate gene expression by controlling and interpreting chromatin modifications. Cancer cells are characterized by altered epigenetic landscapes, and commonly exploit the chromatin regulatory machinery to enforce oncogenic gene expression programs<sup>1</sup>. Although chromatin alterations are, in principle, reversible and often amenable to drug intervention, the promise of targeting such pathways therapeutically has been limited by an incomplete understanding of cancer-specific dependencies on epigenetic regulators. Here we describe a non-biased approach to probe epigenetic vulnerabilities in acute myeloid leukaemia (AML), an aggressive haematopoietic malignancy that is often associated with aberrant chromatin states<sup>2</sup>. By screening a custom library of small hairpin RNAs (shRNAs) targeting known chromatin regulators in a genetically defined AML mouse model, we identify the protein bromodomain-containing 4 (Brd4) as being critically required for disease maintenance. Suppression of Brd4 using shRNAs or the small-molecule inhibitor JQ1 led to robust antileukaemic effects *in vitro* and *in vivo*, accompanied by terminal myeloid differentiation and elimination of leukaemia stem cells. Similar sensitivities were observed in a variety of human AML cell lines and primary patient samples, revealing that JQ1 has broad activity in diverse AML subtypes. The effects of Brd4 suppression are, at least in part, due to its role in sustaining Myc expression to promote aberrant self-renewal, which implicates JQ1 as a pharmacological means to suppress MYC in cancer. Our results establish small-molecule inhibition of Brd4 as a promising therapeutic strategy in AML and, potentially, other cancers, and highlight the utility of RNA interference (RNAi) screening for revealing epigenetic vulnerabilities that can be exploited for direct pharmacological intervention.

AML represents a paradigm for understanding how complex patterns of cooperating genetic and epigenetic alterations lead to tumorigenesis<sup>3,4</sup>. Although this complexity poses a challenge for the development of targeted therapies, diverse gene mutations in AML generally converge functionally to deregulate similar core cellular processes. One key event in AML initiation is the corruption of cell-fate programs to generate leukaemia stem cells that aberrantly self-renew and thereby maintain and propagate the disease<sup>5</sup>. Although it is incompletely understood, this process has been linked to changes in regulatory chromatin modifications<sup>2</sup>. For example, common AML oncogenes such as those encoding the AML1-ETO and MLL fusion proteins induce self-renewal programs at least in part through reprogramming of epigenetic pathways<sup>6,7</sup>. In addition, several genes encoding epigenetic regulators have been identified as targets of somatic mutation in AML<sup>8,9</sup>. Because epigenetic

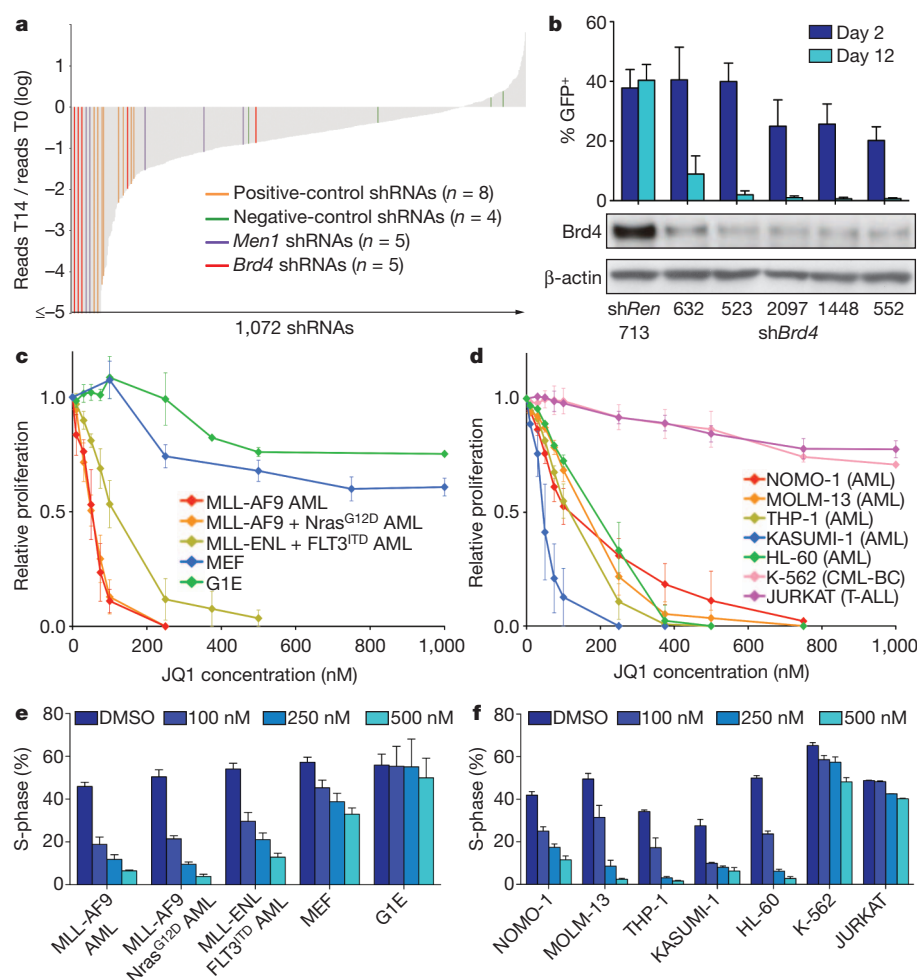
alterations induced by oncogenic stimuli are potentially reversible, chromatin regulators are being explored as candidate drug targets<sup>1</sup>.

To probe epigenetic pathways required for AML maintenance systematically, we built a custom shRNA library targeting 243 known chromatin regulators, including most 'writers', 'readers' and 'erasers' of epigenetic marks (Supplementary Fig. 1 and Supplementary Table 1). This library of 1,094 shRNAs (3–6 per gene) was constructed in TRMPV-Neo, a vector optimized for negative-selection RNAi screening, and was transduced as one pool into an established Tet-on-competent AML mouse model driven by MLL-AF9 and Nras<sup>G12D</sup> (ref. 10). After drug selection, shRNA expression was induced by addition of doxycycline, and changes in library representation after 14 days of culture were monitored using deep sequencing of shRNA guide strands amplified from genomic DNA (Fig. 1a and Supplementary Fig. 2). Using the scoring criterion of more than twenty-fold depletion in each of two independent replicates, 177 shRNAs were strongly depleted. These included all eight positive-control shRNAs targeting essential genes (*Rpa1*, *Rpa3*, *Pcna* and *Polr2b*), as well as several shRNAs targeting two known MLL-AF9 cofactors (*Men1* and *Psip1*)<sup>11,12</sup>. Genes for which at least two independent shRNAs scored were subjected to extensive one-by-one validation using an independent MLL-AF9/Nras<sup>G12D</sup> AML cell line and vector system (Supplementary Fig. 3a). In both the primary screen and validation stages, several shRNAs targeting *Brd4* were among the most strongly depleted, identifying this gene as the top scorer in the screen (Fig. 1a and Supplementary Fig. 3b).

Brd4 is a member of the BET family of bromodomain-containing proteins that bind to acetylated histones to influence transcription<sup>13</sup>. *BRD4* is also a proto-oncogene that can be mutated via chromosomal translocation in a rare form of squamous-cell carcinoma<sup>14</sup>, although a role in leukaemia has not been described. The recent development of small-molecule BET bromodomain inhibitors<sup>15,16</sup>, together with our screening results, prompted us to investigate the suitability of Brd4 as an AML drug target. Five independent *Brd4* shRNAs showed a close correspondence between knockdown efficiency and growth inhibition, indicating on-target effects (Fig. 1b). Suppression of Brd4 led to cell-cycle arrest and apoptosis of leukaemia cells, whereas the equivalent knockdown in immortalized murine embryonic fibroblasts (MEFs) led to only modest cell-cycle inhibition without cytotoxicity (Supplementary Fig. 4a–d). Brd4 knockdown also failed to influence the growth of non-transformed G1E erythroblast cells (Supplementary Fig. 4e). In addition, shRNAs targeting *BRD4* were sufficient to induce cell-cycle arrest in two MLL-AF9<sup>+</sup> human AML lines (Supplementary Fig. 5). Together, these results indicate that Brd4 is a critical requirement in MLL-AF9-induced AML.

<sup>1</sup>Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. <sup>2</sup>Research Institute of Molecular Pathology (IMP), Dr. Bohr-Gasse 7, A-1030 Vienna, Austria. <sup>3</sup>Molecular and Cellular Biology Program, Stony Brook University, Stony Brook, New York 11794, USA. <sup>4</sup>Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. <sup>5</sup>Ludwig Boltzmann Cluster Oncology, Medical University of Vienna, A-1090 Vienna, Austria. <sup>6</sup>Division of Pediatric Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. <sup>7</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. <sup>8</sup>Department of Internal Medicine I, Division of Hematology & Hemostaseology, Medical University of Vienna, A-1090 Vienna, Austria. <sup>9</sup>Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. <sup>10</sup>Department of Laboratory Medicine, University of California at San Francisco, San Francisco, California 94143, USA. <sup>11</sup>Howard Hughes Medical Institute, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

\*These authors contributed equally to this work.



Next, we examined the sensitivity of leukaemia cells to JQ1, a first-in-class small-molecule inhibitor of BET bromodomains with highest affinity for the first bromodomain of Brd4 (ref. 15). Proliferation of mouse MLL-fusion leukaemia cells was notably sensitive to sub-micromolar JQ1 concentrations, as compared to proliferation of fibroblasts and G1E cells (Fig. 1c), correlating with the relative impact of *Brd4* shRNAs on proliferation of these different cell types. We also examined growth-inhibitory effects of JQ1 in a series of established human leukaemia cell lines as well as in adult and paediatric primary leukaemia samples. We observed broad growth-suppressive activity of JQ1 ( $IC_{50} < 500$  nM) in 13 out of 14 AML cell lines (Fig. 1d and Supplementary Fig. 6a, b) and in 12 out of 15 primary human AML samples, representing diverse disease subtypes (Supplementary Fig. 7, 8). In addition, all three primary MLL-rearranged infant leukaemias tested were sensitive to JQ1 (Supplementary Fig. 8), whereas other non-AML-leukaemia and solid-tumour cell lines showed minimal sensitivity to the compound (Fig. 1d and Supplementary Fig. 6c). In all AML lines examined, JQ1 treatment triggered cell-cycle arrest and apoptosis, similar to the effects seen upon shRNA-mediated Brd4 knockdown (Fig. 1e, f and Supplementary Figs 7–9). Together, these data indicate a critical requirement for Brd4 in AML proliferation that can be effectively inhibited using the bromodomain inhibitor JQ1.

We next investigated the relevance of Brd4 to AML progression *in vivo*. To suppress Brd4 in established AML in mice, Tet-on-competent MLL-AF9/Nras<sup>G12D</sup> leukaemia cells were transduced with TRMPV-Neo constructs containing *Brd4* shRNAs or control shRNAs, and transplanted into secondary recipient mice (Supplementary Fig. 10). After disease onset, confirmed by bioluminescent imaging, shRNA expression was induced by doxycycline administration. Subsequent monitoring showed that *Brd4* knockdown resulted in a marked delay

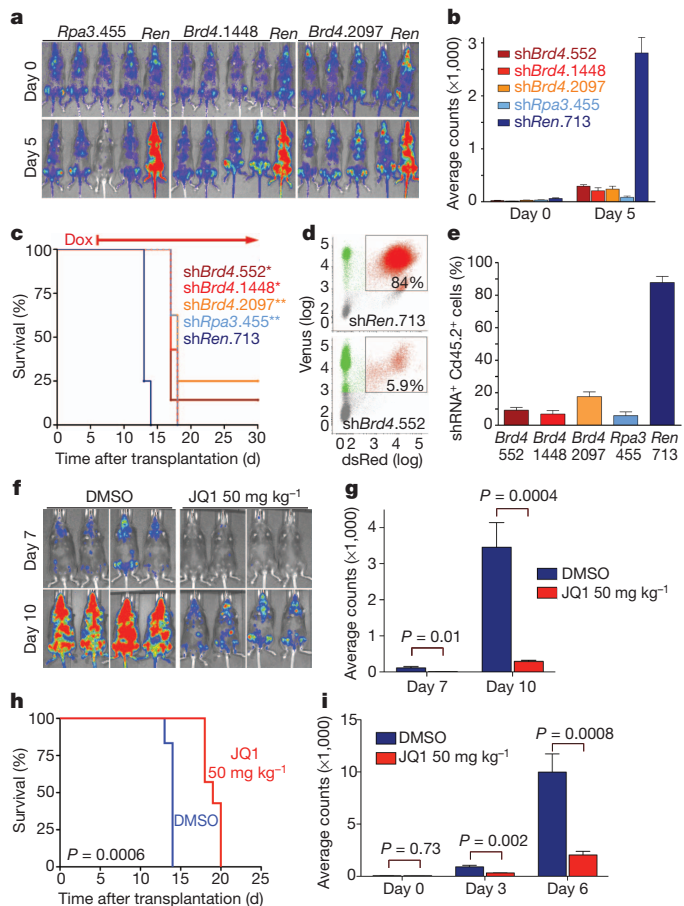
**Figure 1 | AML growth is sensitive to Brd4 inhibition.** **a**, Pooled negative-selection screening in MLL-AF9/Nras<sup>G12D</sup> leukaemia, depicting changes in representation of 1,072 informative shRNAs during 14 days of culture. shRNA abundance ratios were calculated as the number of reads after 14 days of culture on doxycycline (T14) divided by the number of reads before doxycycline treatment (T0), and plotted as the mean of two replicates in ascending order. Completely depleted shRNAs (0 reads at T14,  $n = 71$ ) were plotted as a ratio of  $10^{-5}$ ; highlighted shRNAs in this group are shown with even spacing in alphabetical order. Positive controls included shRNAs targeting *Rpa1*, *Rpa3*, *Pcna* and *Polr2b*. Negative-control shRNAs targeted renilla luciferase or *Braf*. **b**, Percentage of GFP-positive MLL-AF9/Nras<sup>G12D</sup> leukaemia cells 2 days and 12 days after transduction with LMN constructs expressing the indicated shRNAs (top panel). Western blotting of whole-cell lysates prepared from MEF cultures transduced with the indicated TtTMPV-Neo-shRNAs and induced with doxycycline for 5 days (bottom panel). shRen, renilla luciferase shRNA. Representative experiments are shown. **c**, **d**, Proliferation rates of JQ1-treated cells, calculated by measuring the increase in viable cell number after 72 h in culture and fitting data to an exponential growth curve. Results are normalized to the proliferation rate of vehicle/DMSO-treated cells, set to 1 ( $n = 3$ ). CML-BC, chronic myeloid leukaemia blast crisis; T-ALL, T-cell acute lymphoblastic leukaemia. **e**, **f**, Percentage of cells in S-phase (bromodeoxyuridine (BrdU)<sup>+</sup>) after JQ1 treatment for 48 h at the indicated concentrations ( $n = 3$ ). BrdU was pulsed for 30 min in all experiments shown. All error bars represent s.e.m.

in leukaemia progression and a survival benefit (Fig. 2a–c and Supplementary Fig. 11). Taking advantage of the dsRed reporter linked to shRNA expression in the TRMPV-Neo vector<sup>10</sup>, flow-cytometry analysis verified that *Brd4*-shRNA-positive cells were depleted in the terminal leukaemia burden as compared to controls, indicating that the mice succumbed to an outgrowth of *Brd4*-shRNA-negative cells (Fig. 2d, e). Together, these data indicate that RNAi-mediated suppression of Brd4 inhibits leukaemia progression *in vivo*.

To examine whether JQ1 has single-agent activity in AML, mice transplanted with MLL-AF9/Nras<sup>G12D</sup> leukaemia cells were treated either with daily injections of JQ1 (50 mg kg<sup>-1</sup>) or with vehicle. JQ1 administration led to a marked delay in disease progression and significantly extended survival (Fig. 2f–h). JQ1 also showed single-agent activity in an intervention setting, in which treatment was initiated only after disease was detected by bioluminescent imaging (Fig. 2i and Supplementary Fig. 12). Comparable effects were observed in an independent AML mouse model based on expression of AML1-ETO9a and Nras<sup>G12D</sup> and loss of p53 (ref. 17), which is known to be insensitive to conventional chemotherapy (Supplementary Fig. 13). Consistent with previous findings<sup>15</sup>, JQ1 treatment was well tolerated in mice, with little if any impact on normal haematopoiesis (Supplementary Figs 14–16). Collectively, these findings demonstrate that JQ1 has potent and leukaemia-specific effects as a single agent *in vivo*.

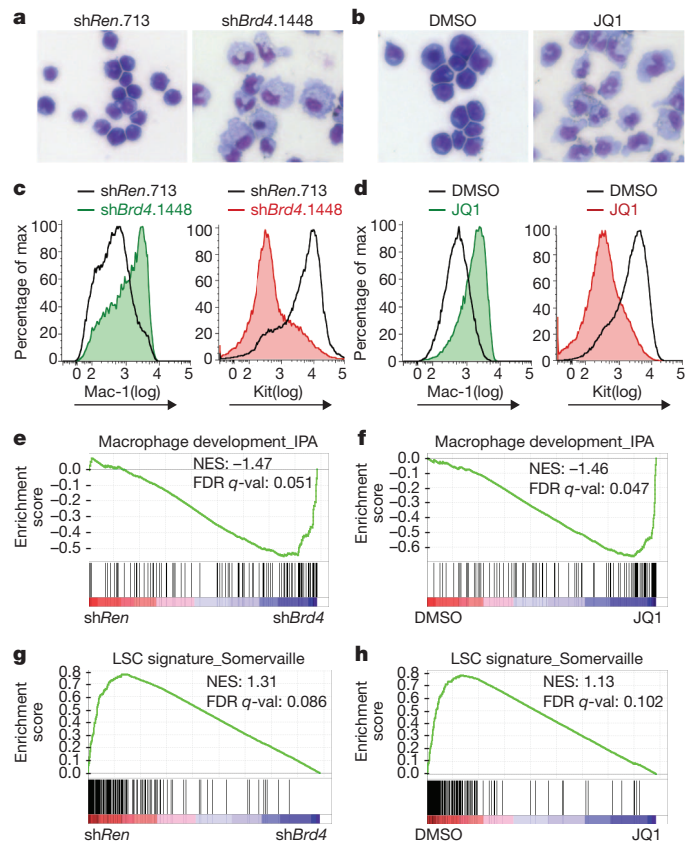
AML is characterized by an expanded self-renewal capacity linked with an inability to complete terminal myeloid differentiation. Therefore, we next considered whether Brd4 influences the differentiation state of leukaemia cells. Both expression of *Brd4* shRNA and JQ1 treatment altered the morphology of MLL-AF9/Nras<sup>G12D</sup> leukaemia from myelomonocytic blasts to cells with a macrophage-like appearance (Fig. 3a, b and Supplementary Fig. 17a). Upon Brd4 inhibition,





**Figure 2 | Brd4 is required for AML progression *in vivo*.** **a**, Bioluminescent imaging of mice transplanted with MLL-AF9/Nras<sup>G12D</sup> leukaemia cells harbouring the indicated TRMPV-Neo-shRNAs. Doxycycline was administered upon disease onset, 6 days after transplant. Day 0 indicates the time of doxycycline treatment. **b**, Quantification of bioluminescent imaging responses after doxycycline treatment. Mean values of four replicate mice are shown. **c**, Kaplan–Meier survival curves of recipient mice transplanted with the indicated TRMPV-Neo-shRNA leukaemia lines. The interval of doxycycline treatment is indicated by the arrow. Each shRNA group contained 6–8 mice. Statistical significance compared to shRen was calculated using a log-rank test; \*,  $P = 0.0001$ ; \*\*,  $P < 0.0001$ . **d**, Representative flow cytometry plots of donor-derived (Cd45.2<sup>+</sup>) bone marrow cells in terminally diseased doxycycline-treated mice. The gate shown includes dsRed<sup>+</sup>/shRNA<sup>+</sup> cells. **e**, Percentage of dsRed<sup>+</sup>/shRNA<sup>+</sup> cells in the Cd45.2<sup>+</sup> terminal leukaemia burden. Mean values of four replicate mice are shown. **f**, Bioluminescent imaging of MLL-AF9/Nras<sup>G12D</sup> leukaemia recipient mice at the indicated day after initiation of treatment with JQ1 (50 mg kg<sup>-1</sup> d<sup>-1</sup>) or DMSO carrier. **g**, Quantification of bioluminescent imaging responses to JQ1 treatment. Mean values of six DMSO- and seven JQ1-treated mice are shown.  $P$  values were calculated using a two-tailed Student's *t*-test. **h**, Kaplan–Meier survival curves of control and JQ1-treated mice. Statistical significance was calculated using a log-rank test. In **f**, **g** and **h**, JQ1 treatment was initiated on day 1 after transplant of 50,000 leukaemia cells. **i**, Quantification of bioluminescent imaging responses to JQ1 treatment in established disease. Treatment of leukaemia mice was initiated 6 days after transplant, when disease first became detectable by imaging. Mean values of six DMSO- and seven JQ1-treated mice are shown.  $P$  values were calculated using a two-tailed Student's *t*-test. All error bars represent s.e.m.

leukaemia cells showed increased surface expression of integrin  $\alpha$ M (Itgam, also known as Mac-1), a myeloid differentiation marker, and decreased expression of Kit, a marker associated with leukaemia stem cells (LSCs) in mouse models of MLL-rearranged leukaemia (Fig. 3c, d and Supplementary Fig. 17b, c)<sup>18,19</sup>. In addition, JQ1 treatment induced morphological signs of maturation phenotypes in most of the primary leukaemia samples tested, albeit to varying degrees (Supplementary Figs 7 and 8).

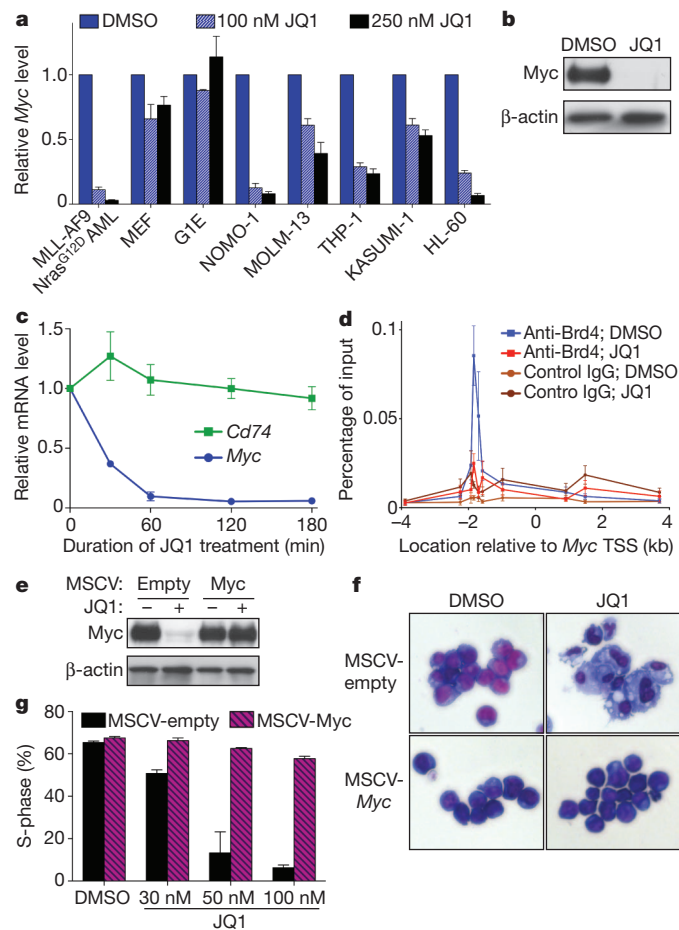


**Figure 3 | Brd4 inhibition leads to myeloid differentiation and leukaemia stem-cell depletion.** **a**, **b**, Light microscopy of May–Grünwald/Giemsa-stained MLL-AF9/Nras<sup>G12D</sup> leukaemia cells after 2 days of doxycycline-induced shRNA expression or 2 days of JQ1 treatment (100 nM). Expression of shRNA was induced in TRMPV-Neo-transduced leukaemia cells. Imaging was performed with a  $\times 40$  objective. Representative images of three biological replicates are shown. **c**, **d**, Flow cytometry analysis of Mac-1 and Kit surface expression after 4 days of shRNA expression or 2 days of JQ1 treatment (100 nM). A representative experiment of three biological replicates is shown. **e**–**h**, GSEA plots evaluating changes in macrophage and LSC gene signatures upon Brd4 inhibition. In **e** and **g**, RNA for expression arrays was obtained from sorted dsRed<sup>+</sup>/shRNA<sup>+</sup> cells (shRen versus three different Brd4 shRNAs) after 2 days of doxycycline induction. In **f** and **h**, microarray data were obtained from leukaemia cells treated for 2 days with DMSO or 100 nM JQ1. NES, normalized enrichment score; FDR  $q$ -val, false discovery rate  $q$ -value (the probability that a gene set with a given NES represents a false-positive finding).

To further investigate whether suppression of Brd4 affects the LSC compartment, we performed gene set enrichment analysis (GSEA) of expression microarray data obtained from Brd4-shRNA-treated and JQ1-treated leukaemia cells<sup>20</sup>. GSEA revealed a marked upregulation of macrophage-specific gene expression after Brd4 inhibition (Fig. 3e, f), as well as global downregulation of a gene expression signature previously shown to discriminate LSCs from non-self-renewing leukaemia cell populations (Fig. 3g, h)<sup>21</sup>. A similar profile of gene expression changes was seen after JQ1 treatment of THP-1 cells, a human MLL-AF9-expressing AML cell line (Supplementary Fig. 18). Although we cannot exclude the involvement of additional cellular targets, the strong concordance between phenotypes induced by Brd4 shRNAs and JQ1 supports Brd4 as the relevant target of JQ1 in AML. Together, these findings indicate that Brd4 is critically required to maintain LSC populations and prevent terminal myeloid differentiation.

Recent evidence indicates that the Myc transcriptional network has an important role in LSC self-renewal<sup>22,23</sup>. Because previous studies also implicate Myc as a potential downstream target of Brd4 (refs 24, 25), we examined whether this regulatory function was relevant to the anti-leukaemic effects of JQ1. Brd4 inhibition with shRNAs or JQ1 led to a

marked reduction of mRNA and protein levels of Myc in MLL-AF9/Nras<sup>G12D</sup> leukaemia, whereas these effects were minimal in MEF and G1E cells (Fig. 4a, b and Supplementary Fig. 19a–c). Downregulation of Myc mRNA levels occurred within 60 min of JQ1 exposure, qualitatively preceding the increased expression of genes related to macrophage differentiation, such as *Cd74* (Fig. 4c). Supporting a direct role in Myc transcriptional regulation, chromatin immunoprecipitation (ChIP) experiments identified a region of focal Brd4 occupancy about 2 kilobases upstream of the Myc promoter and this was eliminated after exposure to competitive JQ1 (Fig. 4d). As expected, RNAi- or JQ1-induced suppression of Brd4 led to a global reduction in expression of



**Figure 4 | JQ1 suppresses the Myc pathway in leukaemia cells.** **a**, Quantitative reverse transcription PCR (qRT-PCR) of relative Myc mRNA levels in the indicated mouse or human cells after a 48 h treatment with indicated JQ1 dose or DMSO. Results were normalized to *Gapdh*, with the relative mRNA level in untreated cells set to 1 ( $n = 3$ ). **b**, Western blotting of whole-cell lysates prepared from MLL-AF9/Nras<sup>G12D</sup> leukaemia cells treated for 48 h with DMSO or 250 nM JQ1. A representative experiment of three biological replicates is shown. **c**, qRT-PCR time course at indicated time points after treatment of MLL-AF9/Nras<sup>G12D</sup> leukaemia cells with 250 nM JQ1. Results were normalized to *Gapdh*, with the relative mRNA level in untreated cells set to 1 ( $n = 3$ ). **d**, ChIP-qPCR performed in MLL-AF9/Nras<sup>G12D</sup> leukaemia cells with the indicated antibodies and PCR primer locations ( $n = 6$  for DMSO;  $n = 4$  for JQ1-treated). TSS, transcription start site. **e**, Western blotting of whole-cell lysates prepared from MLL-AF9/Nras<sup>G12D</sup> leukaemia cells transduced with empty vector or a Myc cDNA-containing MSCV retrovirus. Cells were treated for 48 h with DMSO or 250 nM JQ1. A representative experiment of three biological replicates is shown. **f**, Light microscopy of May-Grünwald/Giemsa-stained MLL-AF9/Nras<sup>G12D</sup> leukaemia cells transduced with an empty vector or with the Myc cDNA. Cells were treated for 5 days with 50 nM JQ1 and imaged using a  $\times 40$  objective. A representative image of three biological replicates is shown. **g**, Quantification of BrdU incorporation after a 30-min pulse in MLL-AF9/Nras<sup>G12D</sup> leukaemia cells transduced with empty control vector or the Myc cDNA. Cells were treated with JQ1 for 5 days at the indicated concentrations ( $n = 3$ ). All error bars shown represent s.e.m.

Myc target genes<sup>22,26</sup> (Supplementary Figs 19d and 20). Notably, JQ1 treatment triggered Myc downregulation in a broad array of mouse and human leukaemia cell lines examined (Fig. 4a, b and Supplementary Fig. 21), indicating that JQ1 may provide a means to suppress the Myc pathway in a range of AML subtypes.

To evaluate whether suppression of Myc confers the growth-inhibitory effects of JQ1, we generated MLL-AF9/Nras<sup>G12D</sup> leukaemia cultures in which the Myc cDNA was ectopically expressed from a retroviral promoter, which resulted in Myc expression levels that were only slightly elevated but entirely resistant to JQ1-induced suppression (Fig. 4e and Supplementary Fig. 22a). Ectopic Myc expression conferred nearly complete resistance to macrophage differentiation and cell-cycle arrest induced by JQ1 and *Brd4* shRNAs (Fig. 4f, g and Supplementary Figs 22b and 23). Furthermore, global gene-expression profiling revealed that most of the JQ1-induced transcriptional changes are probably secondary effects of Myc downregulation (Supplementary Fig. 24). Ectopic Myc expression was unable to prevent JQ1-induced cell death, indicating that Brd4 has Myc-independent roles in regulating cell survival (Supplementary Fig. 22c, d). These findings collectively support a role for Brd4 in maintaining Myc expression to preserve an undifferentiated cellular state in AML.

By taking a non-biased screening approach targeting epigenetic regulators, our study has identified Brd4 as a critical factor required for AML disease maintenance. Because Brd4 is not evidently mutated or overexpressed in AML (Supplementary Fig. 25), the exquisite sensitivity of leukaemia cells to Brd4 inhibition would not have been revealed simply through genetic or transcriptional characterization of this disease. We further show that the bromodomain inhibitor JQ1 has broad activity in diverse AML contexts and, by comparing its effects to those induced by *Brd4* shRNAs, we provide evidence that Brd4 is the primary target for the antileukaemic activity of JQ1. Of note, JQ1 is a first-generation chemical inhibitor yet to be optimized for *in vivo* delivery, with a half-life of only about 1 h in rodents (Supplementary Fig. 26 and ref. 15). The more robust antileukaemic effects seen using *Brd4* shRNAs *in vivo* indicate that second-generation derivatives of this compound may have greater clinical activity. Regardless, our results unambiguously highlight the utility of RNAi screening for revealing candidate drug targets in cancer.

As a competitive inhibitor of the acetyl-lysine binding domain, JQ1 interferes with the ability of Brd4 to 'read' histone acetylation marks that facilitate transcriptional activation<sup>15</sup>. When applied to leukaemia cells, JQ1 interferes with transcriptional circuits supporting self-renewal, thus targeting LSCs and inducing terminal differentiation. In a parallel study, we identified the transcription factor Myb as a critical mediator of addiction to the MLL-AF9 oncogene<sup>31</sup>. Notably, global gene-expression changes observed after genetic or pharmacological inhibition of Brd4 are remarkably similar to those seen upon suppressing Myb<sup>31</sup>, indicating that Myb and Brd4 may intersect functionally in a common transcriptional circuit that is essential for malignant self-renewal. A key downstream effector of both Myb and Brd4 is the oncoprotein Myc (ref. 27), which has been validated as an attractive therapeutic target but has thus far escaped efforts at pharmacological inhibition<sup>28,29</sup>. Although the precise mechanism remains to be further defined, targeting Brd4 abolishes Myc expression and limits self-renewal, with selectivity for the malignant context, thus avoiding the haematopoietic toxicities that may be associated with systemic Myc inhibition<sup>30</sup>. As such, our study may define a general strategy to disarm oncogenic pathways through the direct modulation of the epigenetic machinery.

## METHODS SUMMARY

**Pooled negative-selection RNAi screening.** A customized shRNA library targeting 243 chromatin-regulating mouse genes was designed using miR30-adapted BIOPREDsi predictions, and was generated by PCR-cloning a pool of oligonucleotides synthesized on 55k arrays (Agilent Technologies). Pools of shRNAs were subcloned into the TRMPV-Neo vector (Addgene catalogue no. 27990) together with control shRNAs, and transduced into Tet-on MLL-AF9/Nras<sup>G12D</sup> leukaemia cells for negative-selection screening, essentially as described previously<sup>10</sup>. All



shRNA sequences as well as primary screening data are provided in Supplementary Table 1.

**Animal studies.** All mouse experiments were approved by the Cold Spring Harbor animal care and use committee. For conditional RNAi experiments *in vivo*, Tet-on MLL-AF9/Nras<sup>G12D</sup> leukaemia cells were transduced with TRMPV-Neo-shRNA constructs, followed by transplantation into sub-lethally irradiated recipient mice, as described previously<sup>10</sup>. For shRNA induction, animals were treated with doxycycline in both drinking water (2 mg ml<sup>-1</sup> with 2% sucrose; Sigma-Aldrich) and food (625 mg kg<sup>-1</sup>, Harlan laboratories). For JQ1 treatment trials, a stock of 100 mg ml<sup>-1</sup> JQ1 in DMSO was diluted 20-fold by dropwise addition of a 10% 2-hydroxypropyl- $\beta$ -cyclodextrin carrier (Sigma) under vortexing, yielding a 5 mg ml<sup>-1</sup> final solution. Mice were intraperitoneally injected daily with freshly diluted JQ1 (50 or 100 mg kg<sup>-1</sup>) or a similar volume of carrier containing 5% DMSO.

**Microarray analysis.** Expression microarrays were performed using Affymetrix ST 1.0 GeneChips. Raw microarray data can be accessed at Gene Expression Omnibus, GSE29799. Pathway analysis was performed using GSEA v2.07 software with 1,000 phenotype permutations<sup>20</sup>. All gene sets used for GSEA are provided in Supplementary Table 2.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 March; accepted 23 June 2011.**

**Published online 3 August 2011.**

- Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nature Rev. Cancer* **10**, 457–469 (2010).
- Chen, J., Odenike, O. & Rowley, J. D. Leukaemogenesis: more than mutant genes. *Nature Rev. Cancer* **10**, 23–36 (2010).
- Gilliland, D. G., Jordan, C. T. & Felix, C. A. The molecular basis of leukemia. *Hematology (Am. Soc. Hematol. Educ. Program)* 80–97 (2004).
- Figuerola, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
- Dick, J. E. Stem cell concepts renew cancer research. *Blood* **112**, 4793–4807 (2008).
- Wang, J., Hoshino, T., Redner, R. L., Kajigaya, S. & Liu, J. M. ETO, fusion partner in t(8;21) acute myeloid leukemia, represses transcription by interaction with the human N-CoR/mSin3/HDAC1 complex. *Proc. Natl Acad. Sci. USA* **95**, 10860–10865 (1998).
- Krivtsov, A. V. *et al.* H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* **14**, 355–368 (2008).
- Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
- Ley, T. J. *et al.* *DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
- Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nature Biotechnol.* **29**, 79–83 (2011).
- Yokoyama, A. & Cleary, M. L. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* **14**, 36–46 (2008).
- Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–218 (2005).
- Wu, S. Y. & Chiang, C. M. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J. Biol. Chem.* **282**, 13141–13145 (2007).
- French, C. A. *et al.* *BRD4-NUT* fusion oncogene: a novel mechanism in aggressive carcinoma. *Cancer Res.* **63**, 304–307 (2003).
- Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
- Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).
- Zuber, J. *et al.* Mouse models of human AML accurately predict chemotherapy response. *Genes Dev.* **23**, 877–889 (2009).
- Somervaille, T. C. & Cleary, M. L. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell* **10**, 257–268 (2006).
- Krivtsov, A. V. *et al.* Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Somervaille, T. C. *et al.* Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* **4**, 129–140 (2009).
- Kim, J. *et al.* A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313–324 (2010).
- Wong, P. *et al.* The miR-17-92 microRNA polycistron regulates MLL leukemia stem cell potential by modulating p21 expression. *Cancer Res.* **70**, 3833–3842 (2010).
- Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005).
- Yang, Z., He, N. & Zhou, Q. Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. *Mol. Cell. Biol.* **28**, 967–976 (2008).
- Schuhmacher, M. *et al.* The transcriptional program of a human B cell line in response to Myc. *Nucleic Acids Res.* **29**, 397–406 (2001).
- Schmidt, M., Nazarov, V., Stevens, L., Watson, R. & Wolff, L. Regulation of the resident chromosomal copy of c-myc by c-Myb is involved in myeloid leukemogenesis. *Mol. Cell. Biol.* **20**, 1970–1981 (2000).
- Soucek, L. *et al.* Modelling Myc inhibition as a cancer therapy. *Nature* **455**, 679–683 (2008).
- Felsher, D. W. & Bishop, J. M. Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell* **4**, 199–207 (1999).
- Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
- Zuber, J. *et al.* An integrated approach to dissecting oncogene addiction implicates a Myb coordinated self-renewal program as essential for leukemia maintenance. *Genes Dev.* doi:10.1101/gad.1726911 (in the press).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Ma, S. Muller and M. Weissenboeck for technical assistance; J. Simon, E. Earl and L. Bianco for support with mouse work; C. dos Santos for assistance with LSK FACS analysis; S. Hearn for microscopy support; G. Hannon, K. Chang, and E. Hodges for shRNA technology support; A. Gordon and M. Hammell for bioinformatics support; L. Dow for assistance with mouse pathology sample preparation; and G. Blobel for comments on the manuscript. We thank the Don Monti Memorial Research Foundation and Laurie Strauss Leukemia Foundation for research support. J.Z. was supported by a research fellowship from the German Research Foundation (DFG) and by the Andrew Seligson Memorial Clinical Fellowship at CSHL; A.R.R. was supported by an NIH traineeship and the Barbara McClintock fellowship; J.E.B. and J.Q. are supported by the Damon-Runyon Cancer Research Foundation and Smith Family Foundation. P.B. is supported by a Damon Runyon-Lilly Clinical Investigator Award and a Leukemia and Lymphoma Society (LLS) Translational Research Program Grant, and is an LLS Scholar in Clinical Research. S.W.L. is supported by a Specialized Center of Research (SCOR) grant from the Leukemia and Lymphoma Society of America, a Cancer Target Discovery and Development (CTD2) grant from the National Cancer Institute, and by the Howard Hughes Medical Institute; C.R.V., J.S. and E.W. are supported by the CSHL President's Council and the SASS Foundation for Medical Research.

**Author Contributions** J.Z., J.S. and C.R.V. designed research, performed experiments and analysed data. E.W., A.R.R. and M.J.T. performed experiments and analysed data. C.J. performed microarray experiments. A.C. assisted with shRNA library design. H.H., E.A.S., D.M., K.B., P.B. and P.V. performed experiments with primary leukaemia specimens. J.C.M. and M.W. generated and provided engineered AML lines. S.C.K. performed histological analysis. J.Q. and J.E.B. designed research and synthesized and supplied JQ1. J.Z., S.W.L. and C.R.V. co-wrote the paper. S.W.L. and C.R.V. supervised the research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Raw microarray data can be accessed at Gene Expression Omnibus, GSE29799. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.R.V. ([vakoc@cshl.edu](mailto:vakoc@cshl.edu)) or S.W.L. ([lowe@cshl.edu](mailto:lowe@cshl.edu)).



## METHODS

**Pooled negative-selection RNAi screening.** A custom shRNA library targeting 243 chromatin-regulating mouse genes was designed using miR30-adapted BIPREDISI predictions (six shRNAs per gene) and constructed by PCR-cloning a pool of oligonucleotides synthesized on 55k customized arrays (Agilent Technologies) as previously described<sup>10</sup>. After sequence verification, 1,094 shRNAs (3–6 per gene) were combined with several positive- and negative-control shRNAs at equal concentrations in one pool. This pool was subcloned into TRMPV-Neo and transduced into Tet-on MLL-AF9/Nras<sup>G12D</sup> leukaemia cells using conditions that predominantly lead to a single retroviral integration and represent each shRNA in a calculated number of >500 cells (a total of thirty million cells at infection, 2% transduction efficiency). Transduced cells were selected for 5 days using 1 mg ml<sup>-1</sup> G418 (Invitrogen); at each passage more than twenty million cells were maintained to preserve library representation throughout the experiment. After drug selection, T0 samples were obtained (~twenty million cells per replicate) and cells were subsequently cultured with 0.5 mg ml<sup>-1</sup> G418 and 1 µg ml<sup>-1</sup> doxycycline to induce shRNA expression. After 14 days (12 passages, T14), about fifteen million shRNA-expressing (dsRed<sup>+</sup>/Venus<sup>+</sup>) cells were sorted for each replicate using a FACSAriaII (BD Biosciences). Genomic DNA from T0 and T14 samples was isolated by two rounds of phenol extraction using PhaseLock tubes (5prime) followed by isopropanol precipitation. Deep-sequencing template libraries were generated by PCR amplification of shRNA guide strands as previously described<sup>10</sup>. Libraries were analysed on an Illumina Genome Analyser at a final concentration of 8 pM; 18 nucleotides of the guide strand were sequenced using a custom primer (miR30EcoRISeg, TAGCCCTTGAATTCGAGGCAGTAGGCA). To provide a sufficient baseline for detecting shRNA depletion in experimental samples, we aimed to acquire >500 reads per shRNA in the T0 sample, which required more than ten million reads per sample to compensate for disparities in shRNA representation inherent in the pooled plasmid preparation or introduced by PCR biases. With these conditions, we acquired T0 baselines of >500 reads for 1,072 (97% of all) shRNAs. Sequence processing was performed using a customized Galaxy platform<sup>32</sup>. For each shRNA and condition, the number of matching reads was normalized to the total number of library-specific reads per lane and imported into a database for further analysis (Access 2003, Microsoft). All shRNA sequences are provided in Supplementary Table 1.

**Animal studies.** All mouse experiments were approved by the Cold Spring Harbor animal care and use committee. Leukaemia cells were transplanted by tail-vein injection of  $1 \times 10^6$  cells into sub-lethally (5.5 Gy) irradiated B6/SJL(CD45.1) recipient mice. For whole-body bioluminescent imaging, mice were intraperitoneally injected with 50 mg kg<sup>-1</sup> D-Luciferin (Goldbio), and after 10 min, analysed using an IVIS Spectrum system (Caliper LifeSciences). Quantification was performed using Living Image software (Caliper LifeSciences) with standardized rectangular regions of interests covering the mouse trunk and extremities. For shRNA induction, animals were treated with doxycycline in both drinking water (2 mg ml<sup>-1</sup> with 2% sucrose; Sigma-Aldrich) and food (625 mg kg<sup>-1</sup>, Harlan laboratories). For JQ1 treatment trials, a stock of 100 mg ml<sup>-1</sup> JQ1 in DMSO was 20-fold diluted by dropwise addition of a 10% 2-hydroxypropyl-β-cyclodextrin carrier (Sigma) under vortexing, yielding a 5 mg ml<sup>-1</sup> final solution. Mice were intraperitoneally injected daily with freshly diluted JQ1 (50 or 100 mg kg<sup>-1</sup>) or a similar volume of carrier containing 5% DMSO.

**Plasmids.** For conditional RNAi experiments, shRNAs were expressed from either the TRMPV-Neo vector or the TtTMPV-Neo vector, which have been described previously (and are available as Addgene catalogue nos 27990 and 27993)<sup>10</sup>. For screen validation, shRNAs were cloned into LMN(MSCV-miR30-PGK-NeoR-IRES-GFP), which was generated from LMP<sup>33</sup> by replacing the PuroR transgene with a NeoR cassette. For Myc rescue experiments, the wild-type mouse Myc cDNA was subcloned into MSCV-PGK-Puro-IRES-GFP (MSCV-PIG)<sup>34</sup>.

**Cell culture.** All mouse MLL-leukaemia cell lines were derived from bone marrow obtained from terminally ill recipient mice, and were cultured in RPMI 1640 (Gibco-Invitrogen) supplemented with 10% FBS, 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin. MLL-AF9(alone), MLL-AF9/Nras<sup>G12D</sup>, Tet-on MLL-AF9/Nras<sup>G12D</sup> and MLL-ENL/FLT3<sup>ITD</sup> cell cultures were derived as described previously<sup>10,17</sup>. Tet-on immortalized MEF cultures were described previously<sup>10</sup>. G1E cells were provided by M. Weiss. MEF cells were grown in DMEM with 10% FBS, 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin and 1% glutamine (GIBCO). G1E cells were grown in IMDM with 15% FBS, 100 U ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin, 2 U ml<sup>-1</sup> erythropoietin (Sigma) and 10% Kit-ligand-conditioned medium. All human leukaemia cell lines were cultured in RPMI-1640 with 10% FBS, 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin, except KASUMI-1 cells, which were cultured in 20% FBS. NOMO-1, MOLM-13, EoL-1, NB4, HNT-34 and CMK were purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). KASUMI-1, HL-60, MV4-11, KG1, HEL, THP-1,

B16-F10 and IMR-90 were obtained from ATCC. K-562 cells were provided by M. Carroll. U2OS, HeLa and Jurkat were provided by the CSHL tissue culture facility. CD34.MA9.NRAS and CD34.MA9.FLT3 cells were generated by retroviral transduction of umbilical-cord-blood CD34<sup>+</sup> cells as described previously<sup>35,36</sup>. All retroviral packaging was performed using ecotropic Phoenix cells according to established protocols ([http://www.stanford.edu/group/nolan/tutorials/retpkg\\_1\\_packlines.html](http://www.stanford.edu/group/nolan/tutorials/retpkg_1_packlines.html)).

**Western blotting.** For Brd4 immunoblotting, 30 µg of whole-cell lysate RIPA extracts (25 mM Tris (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) were loaded into each lane. For Myc immunoblotting, cells were lysed directly in Laemmli buffer and about 50,000 cell-equivalents were loaded into each lane. Protein extracts were resolved by SDS polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose for blotting.

**Proliferation assay.** Competitive proliferation assays using shRNAs in LMN or TRMPV-/TtTMPV-Neo vectors were performed as outlined in Supplementary Fig. 3a and as described previously<sup>10</sup>, respectively. Proliferation assays for JQ1 *in vitro* testing were performed by counting the increase in viable cell numbers over 72 h in the presence of different JQ1 concentrations. Dead cells were excluded using propidium iodide (PI) staining. Measurements of cell concentration were performed on a Guava EasyCyte (Millipore), gating only viable cells (FSC/SSC/PI<sup>-</sup>). Proliferation rates were calculated using the equation  $\ln(\text{cell concentration at } 72 \text{ h}) / \ln(\text{cell concentration at } 0 \text{ h}) / 72$ . Relative proliferation rates were calculated by normalizing to the rate of DMSO-treated cells.

**May-Grünwald-Giemsa cytospin staining.** MLL-AF9/Nras<sup>G12D</sup> leukaemia cells were treated with 1 µg ml<sup>-1</sup> doxycycline for 2 days to induce shRNA expression from TRMPV-Neo or TtTMPV vectors, or treated with 100 nM JQ1 for 2 days. 50,000 cells were resuspended in 100 µl FACS buffer (5% FBS, 0.05% NaN<sub>3</sub> in PBS) and cytospun onto glass slides using a Shandon Cytospin 2 Centrifuge at 500 rpm for 5 min. May-Grünwald (Sigma) and Giemsa (Sigma) stainings were performed according to manufacturer's protocols. Images were collected using a Zeiss Observer Microscope with a ×40 objective.

**BrdU cell-cycle analysis and annexin V flow cytometry.** BrdU incorporation assays were performed according to the manufacturer's protocol (BD, APC BrdU flow kit), with cells pulsed with BrdU for 30 min. Cells were co-stained with 7-aminoactinomycin D or 4',6-diamidino-2-phenylindole (DAPI) for DNA content measurement. For all conditional shRNA experiments, the analysis was gated on Venus<sup>+</sup>/dsRed<sup>+</sup> (shRNA<sup>+</sup>) cell populations. Annexin V apoptosis staining was performed according to the manufacturer's protocol (BD, APC annexin V). To analyse shRNA-mediated induction of apoptosis specifically, annexin V was quantified in viable shRNA-expressing cells (FSC/SSC; Venus<sup>+</sup>/dsRed<sup>+</sup>). Notably, this gating selectively evaluates early apoptotic cells (Annexin V<sup>+</sup>, DAPI<sup>-</sup>), excluding accumulated dead cells (Annexin V<sup>+</sup>, DAPI<sup>+</sup>). All analyses were performed using FlowJo software (Tree Star).

**shRNA experiments in human AML cell lines.** THP-1 and MOLM-13 cells were modified to express the ecotropic receptor and rtTA3 using retroviral transduction of MSCV-RIEP (MSCV-rtTA3-IRES-EcoR-PGK-Puro) followed by drug selection (0.5 and 1 µg ml<sup>-1</sup> puromycin for 1 week, respectively). The resulting cell lines were transduced with ecotropically packaged TRMPV-Neo-shRNA retroviruses, selected with 400 µg ml<sup>-1</sup> G418 for 1 week and treated with 1 µg ml<sup>-1</sup> doxycycline to induce shRNA expression. The relative change in Venus<sup>+</sup>/dsRed<sup>+</sup> (shRNA<sup>+</sup>) cells was monitored on a Guava EasyCyte (Millipore). BrdU cell-cycle analysis was performed as described above.

**Adult primary leukaemia sample analysis.** The study was approved by the Institutional Review Board (ethics committee) of the Medical University of Vienna. Primary leukaemic cells were obtained from peripheral blood or bone marrow aspirate samples. Informed consent was obtained before blood donation or bone marrow puncture in each case. Diagnoses were established according to criteria provided by the French-American-British (FAB) cooperative study group<sup>37,38</sup> and the World Health Organization (WHO)<sup>39</sup>. Mononuclear cells were prepared using Ficoll and stored in liquid nitrogen until used. HL60 and MOLM13 cell lines (obtained from DSMZ) were included as controls. After thawing, the viability of AML cells ranged from 70% to 99% as assessed by trypan blue exclusion. Primary cells (thawed mononuclear cells, 5–10 × 10<sup>4</sup> cells per well) and cell lines (1–5 × 10<sup>4</sup> cells per well) were cultured in 96-well microtitre plates (TPP) in RPMI-1640 medium (PAA laboratories) with 10% fetal calf serum (FCS, Pasching) in the absence or presence of JQ1 (10–5,000 nM) at 37 °C (5% CO<sub>2</sub>) for 48 h. In selected experiments, primary AML cells were incubated with JQ1 in the presence or absence of a cocktail of proliferation-inducing cytokines: recombinant human (rh) G-CSF, 100 ng ml<sup>-1</sup> (Amgen), rhSCF, 100 ng ml<sup>-1</sup> (Peprotech) and rhIL-3, 100 ng ml<sup>-1</sup> (Novartis). After 48 h, 0.5 µCi <sup>3</sup>H-thymidine was added (16 h). Cells were then harvested on filter membranes in a Filtermate 196 harvester (Packard Bioscience). Filters were air-dried and the bound radioactivity was measured in a β-counter (Top-Count NXT, Packard Bioscience). All experiments

were performed in triplicates. Proliferation was calculated as a percentage of control (cells kept in control medium), and the inhibitory effects of JQ1 were expressed as IC<sub>50</sub> values. In 7 out of 12 patients, drug-exposed cells were analysed for morphologic signs of differentiation by Wright–Giemsa staining on cytospin slides. The thymidine incorporation assay was chosen as a proliferation assay because of its superior sensitivity and ease of implementation for suspension cells, as compared to other proliferation assays, such as MTT.

**Paediatric primary leukaemia sample analysis.** Diagnostic bone marrow samples were collected, under protocols approved by an institutional review board, from newly diagnosed children with acute leukaemia. Informed consent was obtained in accordance with the Helsinki protocol. At the time of collection, primary leukaemic cells were enriched by density centrifugation using Ficoll-Paque PLUS (GE Healthcare) and subsequently stored in liquid nitrogen. Vials of cryopreserved cells were thawed, resuspended in media, and live leukaemic cells were enriched by density centrifugation. Cells were maintained in supplemented media with 20% FBS. All leukaemia cell cultures were incubated at 37 °C in 5% CO<sub>2</sub>. Primary leukaemia samples were treated with dose-ranges of JQ1 and vehicle control for 72 h in 96-well plates. For the annexin binding assays, cells were harvested and stained with Annexin V-PE and 7-AAD (BD Pharmingen), read on a FACSCalibur and analysed with FlowJo software (Tree Star). For the WST-1 assays, WST-1 reagent (Roche Diagnostics) was added to the culture medium (1:10 dilution) and absorbance was measured at 450 nm using a Bio-Rad model 680 microplate reader (Bio-Rad Laboratories). WST-1 assays were performed in triplicate. Primary leukaemia samples were treated with 250 nM JQ1 and vehicle control for 48 h in 96-well plates. Cytospins were prepared at baseline, 24 h and 48 h and stained with Wright–Giemsa solution (Sigma-Aldrich). Images were acquired using a Nikon Eclipse E600 microscope system (Nikon). Although similar to other metabolic assays measuring cell proliferation (for example, MTT), WST-1 has superior sensitivity for the assessment of cytotoxicity in primary leukaemia samples.

**Histological analysis of bone marrow.** Paraffin-embedded sections were stained with haematoxylin and eosin (H&E). Photographs were taken on a Nikon Eclipse 80i microscope with a Nikon Digital Sight camera using NIS-Elements F2.30 software at a resolution of 2560 × 1920. Using Adobe Photoshop CS2, images were re-sized and set at a resolution of 300 pixels inch<sup>-1</sup>, autocontrast was applied and unsharp mask was used to improve image clarity.

**FACS evaluation of normal haematopoiesis.** Bone marrow cells were obtained by flushing mouse femurs and tibias, followed by erythrocyte lysis with ACK buffer (150 mM NH<sub>4</sub>Cl, 10 mM KHCO<sub>3</sub> and 0.1 mM EDTA). Samples were washed in FACS buffer (5% FBS, 0.05% NaN<sub>3</sub> in PBS), followed by staining (two million cells in 100 µl of FACS buffer) for 1 h. Antibody dilutions used were: mouse haematopoietic lineage eFluor 450 cocktail (1:100), PE-Cy7 anti-mouse Kit (1:50), APC anti-mouse Sca-1 (1:100), APC anti-mouse B220 (1:100), APC anti-mouse Cd11b (1:100), APC anti-mouse TER-119 (1:100) and APC anti-mouse Gr-1 (1:100). Stained samples were analysed on an LSRII flow cytometer. Data analysis was performed using FlowJo software (Treestar).

**Expression microarrays.** Microarrays were performed through the CSHL microarray shared resource. RNA was isolated from 10<sup>7</sup> cells using RNeasy Mini Kit (Qiagen). RNA quality was assessed on an Agilent 2100 Bioanalyser, RNA 6000 Pico Series II Chips (Agilent) and samples with a RIN score of 7.0 or greater passed. RNA was amplified by a modified Eberwine technique, amplified antisense RNA was then converted to cDNA using a WT Expression kit (Ambion). The cDNA was then fragmented and terminally labelled with biotin, using the Affymetrix GeneChip WT Terminal Labelling kit (Affymetrix). Samples were then prepared for hybridization, hybridized, washed and scanned according to the manufacturer's instructions on Mouse Gene ST 1.0 GeneChips (Affymetrix). Affymetrix Expression Console QC metrics were used to pass the image data. Raw data was processed by Affymetrix and Limma packages in R-based Bioconductor. Heatmaps were generated using GenePattern software<sup>40</sup>; RMA-processed microarray data was converted into a log<sub>2</sub> scale, selected gene lists were row-normalized and visualized using the HeatMapImage module on GenePattern. All raw microarray data files are available from the Gene Expression Omnibus (GSE29799).

**GSEA analysis.** Gene set enrichment analysis<sup>20</sup> was performed using GSEA v2.07 software with 1,000 phenotype permutations. Leukaemia-stem-cell and *Myc* gene sets were obtained from the indicated publications<sup>21,22,26</sup>. The macrophage-development gene set was obtained from the Ingenuity Pathway Analysis (IPA) software (Ingenuity). To perform GSEA on human microarray data, mouse gene sets were converted into human gene names using bioDBNet dbWalk (<http://biobdnet.abcc.ncicrf.gov/db/dbWalk.php>) or manually using the NCBI database. A detailed description of GSEA methodology and interpretation is provided at <http://www.broadinstitute.org/gsea/doc/GSEASUserGuideFrame.html>. In brief, the normalized enrichment score (NES) provides 'the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes'. The false discovery rate *q*-value (FDR *q*-val) is 'the estimated probability that a gene set with a given NES represents a false positive finding'. In general, given the lack of coherence in most expression datasets and the relatively small number of gene sets being analyzed, an FDR cutoff of 25% is appropriate. Gene sets used in this study are included in Supplementary Table 2.

**Chromatin immunoprecipitation.** ChIP assays were performed exactly as described<sup>41</sup>. Crosslinking was performed with sequential EGS (Pierce) and formaldehyde<sup>42</sup>. All results were quantified by quantitative PCR performed using SYBR green (ABI) on an ABI 7900HT. Each immunoprecipitate signal was referenced to an input standard-curve dilution series (immunoprecipitate/input) to normalize for differences in starting cell number and for primer amplification efficiency.

**qRT-PCR.** RNA was prepared using TRIzol reagent (Invitrogen). Synthesis of cDNA was performed using qScript cDNA SuperMix (Quanta Biosciences). Quantitative PCR analysis was performed on an ABI 7900HT with SYBR green (ABI). All signals were quantified using the ΔCt method. All signals were normalized to the levels of *Gapdh*.

**Antibodies.** The anti-Brd4 antibody used for western blotting was a gift from G. Blobel. The anti-Brd4 antibody used for ChIP was purchased from Sigma (HPA015055). The anti-Myc antibody was purchased from Epitomics (1472-1). Antibodies used in FACS were: APC anti-mouse CD117/Kit (Biolegend, 105811), APC anti-mouse CD11b (Biolegend, 101211), Pacific Blue anti-mouse CD45.2 (Biolegend, 109820), mouse haematopoietic lineage eFluor 450 cocktail (eBioscience, 88-7772-72), APC anti-mouse CD45R/B220 (Biolegend, 103212), APC anti-mouse TER-119/erythroid cells (Biolegend, 116212), APC anti-mouse Ly-6G/Gr-1 (eBioscience, 17-5931), PE-Cy7 anti-mouse CD117/Kit (eBioscience, 25-1171-82) and APC anti-mouse Sca-1 (eBioscience, 17-5981-81). The anti-β-actin HRP antibody was purchased from Sigma (A3854).

32. Taylor, J., Schenck, I., Blankenberg, D. & Nekrutenko, A. Using galaxy to perform large-scale interactive data analyses in *Curr. Protoc. Bioinformatics* Ch. 10, Unit 10.5 (2007).
33. Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
34. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nature Genet.* **33**, 396–400 (2003).
35. Wei, J. *et al.* Microenvironment determines lineage fate in a human model of MLL-AF9 leukemia. *Cancer Cell* **13**, 483–495 (2008).
36. Wunderlich, M. & Mulloy, J. C. Model systems for examining effects of leukemia-associated oncogenes in primary human CD34<sup>+</sup> cells via retroviral transduction. *Methods Mol. Biol.* **538**, 263–285 (2009).
37. Bennett, J. M. *et al.* Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br. J. Haematol.* **33**, 451–458 (1976).
38. Bennett, J. M. *et al.* Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British cooperative group. *Ann. Intern. Med.* **103**, 620–625 (1985).
39. Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937–951 (2009).
40. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
41. Steger, D. J. *et al.* DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol. Cell. Biol.* **28**, 2825–2839 (2008).
42. Zeng, P. Y., Vakoc, C. R., Chen, Z. C., Blobel, G. A. & Berger, S. L. *In vivo* dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* **41**, 694–698 (2006).

## CORRIGENDUM

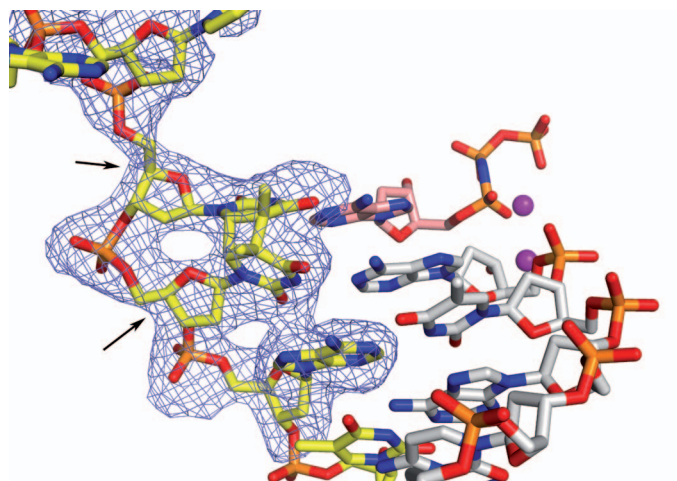
doi:10.1038/nature10338

### Structure and mechanism of human DNA polymerase $\eta$

Christian Biertümpfel, Ye Zhao, Yuji Kondo, Santiago Ramón-Maiques, Mark Gregory, Jae Young Lee, Chikahide Masutani, Alan R. Lehmann, Fumio Hanaoka & Wei Yang

*Nature* **465**, 1044–1048 (2010)

It was brought to our attention by D. Guillaume (Chimie Thérapeutique, CNRS) that deoxyribose C4' of the cyclobutane pyrimidine dimer (CPD) photoproduct in the TT2 structure (Protein Data Bank accession number PDB 3MR4) has an inverted configuration. After further refinement, we obtained the correct sugar configuration (see Fig. 1 below). A new accession number (PDB 3SI8) has been given to the corrected structure.



**Figure 1** | The template (yellow), primer (grey) and incoming nucleotide (pink) are shown as sticks. The omit map of the thymine dimer and surrounding nucleotides is shown as a blue mesh. Arrowheads indicate the C4' of the crosslinked bases.